AWARD NUMBER: W81XWH-13-1-0449

TITLE: Differential splicing of oncogenes and tumor suppressor genes in African and Caucasian American populations: contributing factor in prostate cancer disparities?

PRINCIPAL INVESTIGATOR: Norman H Lee, PhD

CONTRACTING ORGANIZATION: George Washington University School of Medicine
Washington, DC 20052

REPORT DATE: December 2017

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2017	Final	30Sep2013 - 29Sep2017

4. TITLE AND SUBTITLE
Differential splicing of oncogenes and tumor suppressor genes in African and Caucasian American populations: contributing factor in prostate cancer disparities?

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-13-1-0449

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Norman H Lee, PhD; Bi-Dar Wang, PhD; Jacqueline Olender (PhD graduate student)

E-Mail: nhlee@gwu.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
George Washington University School of Medicine
2300 I Street, NW
Ross Hall
Washington, DC 20037

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The overarching goal of this grant is to characterize differential splicing of oncogenes in African American (AA) versus Caucasian American (CA) prostate cancer (PCa). We focused our efforts on two oncogenes, phosphatidylinositol-4,5-bisphosphate 3-kinase delta (*PIK3CD*) and fibroblast growth factor receptor 3 (*FGFR3*). We cloned novel short (-*S*) splice variants *PIK3CD-S* (missing exon 20 due to exon skipping event) and *FGFR3-S* (missing exon 14) that are enriched in AA PCa specimens. PCa cell lines ectopically over-expressing AA-enriched *PIK3CD-S* exhibited enhanced activation of the PI3K/AKT pathway compared to the same lines over-expressing the CA-enriched long (-*L*) variant *PIK3CD-L* (retains exon 20). Moreover, proliferative capacity of the CA-variant lines was sensitive to inhibition by CAL-101, a small molecule inhibitor designed specifically against PIK3CD. In contrast, proliferative capacity of the AA-variant lines was resistant to CAL-101 inhibition. And these findings (CA variants sensitive and AA variants insensitive to CAL-101) were recapitulated in a xenograft mouse model of proliferation and metastasis. Analogously, we demonstrate that *FGFR3-S*: i) encodes a more aggressive oncogenic signaling protein compared to CA-enriched *FGFR3-L* (retains exon 14) as defined by *in vitro* assays, ii) is associated with worse prognosis in patients, and iii) is resistant to the tyrosine kinase inhibitor dovitinib (potential treatment for metastatic castrate-resistant PCa). Our discovery portends a genetic screening test for aggressive tumors that are resistant to small molecule inhibitors.

15. SUBJECT TERMS
prostate cancer, cancer health disparities, alternative splicing, African American, European American, oncogenes, tumor suppressor genes

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	35	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. Introduction……. 1

2. Keywords……. 1

3. Accomplishments……...……... 1

4. Impact……...…… 15

5. Changes/Problems...…..…… 16

6. Products…….……..…..……. 16

7. Participants & Other Collaborating Organizations…… 17

8. Special Reporting Requirements…… 18

9. Appendices…… 19

1. INTRODUCTION

There are striking population/race disparities in prostate cancer (PCa) risk and survival outcome borne out of current health statistics data. This is particularly evident between African Americans (AA) and their European American (EA) counterparts. Epidemiologic studies have shown that higher mortality and recurrence rates for prostate cancer are still evident in AA men even after adjustment for socioeconomic status, environmental factors and health care access. Thus, it is likely that intrinsic biological differences account for some of the cancer disparities. Our overarching hypothesis is that the biological component of prostate cancer health disparities is due, in part, to population-dependent differential splicing of oncogenes and tumor suppressor genes in cancer specimens. The application of genomic approaches has identified splice variants in AA specimens, but absent in EA specimens, encoding more aggressive oncogenic proteins, thereby producing a more cancerous phenotype.

2. KEYWORDS

Prostate cancer, cancer health disparities, alternative RNA splicing, African American, European American, oncogenes, tumor suppressor genes, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta, fibroblast growth factor receptor 3

3. ACCOMPLISHMENTS

Goals as stated in SOW:

Specific Aim 1. To define splice variant pairs (AA-specific variant versus EA-counterpart variant) associated with differential oncogenic behavior *in vitro*, and to delineate the mechanism of action.

Task 1. **Full-length cloning and *in vitro* validation of splice variant pairs**. Subtasks will be run concurrently and are as follows:

1a. Full-length cloning of splice variant pairs and ectopic over-expression into PCa cell lines.

1b. *In vitro* validation of differential oncogenic behavior by full-length splice variant pairs. Splice variant pairs (e.g. AA-specific versus EA-counterpart variant of PIK3CD and FGFR3) will be individually over-expressed in the same PCa cell line background, and screened for differential oncogenic behavior.

1c. *In vitro* validation of differential protein/enzyme activity by full-length splice variant pairs. Splice variant pairs will be individually over-expressed into appropriate cell line for enzyme activity assays and/or assessment of downstream activation of cell signaling components. Activation of downstream signaling components by splice variants will be assessed, for example, by measuring phosphorylation of downstream signaling components with phospho-specific antibodies (e.g. phospho-Akt, phospho-ERK, etc.).

Task 2. ***In vitro* screening and full-length cloning of additional splice variant pairs**. Subtasks will be run concurrently and are as follows:

2a. Exon-targeting and splice junction-targeting siRNAs will be used in appropriate PCa cell lines to identify splice variant pairs exhibiting differential oncogenic behavior following knockdown.

2b. From subtask 2a, we will select 5-10 splice variant pairs that exhibited differential oncogenic behavior for full-length cloning and ectopic over-expression in appropriate cell lines.

2c. Cell lines over-expressing individual full-length variant pairs (e.g. AA-specific variant versus EA-counterpart variant) will be validated *in vitro* for differential oncogenic behavior using *in vitro* screens described in subtask 1b. We will also test for differential sensitivity of splice variant pairs to small molecule inhibitors, if available.

2d. Cell lines over-expressing individual variant pairs (e.g. AA-specific variant versus EA-counterpart variant) will be screened *in vitro* for differential protein/enzyme activity and cell signaling as described in subtask 1c. We will also test for differential sensitivity of splice variant pairs to small molecule inhibitors, if available.

Specific Aim 2. To characterize oncogenic differences of splice variant pairs *in vivo* using xenograft animal models.

Task 1. **Validate differential oncogenic behavior of the splice variant pair for PIK3CD *in vivo***. Stably expressed S (AA-specific) or L variants (EA-counterpart) of *PIK3CD* in appropriate cell line(s) will be transplanted ($1x10^6$ to 10^7 cells) into male SCID-NOD immuno-deficient mice for proliferation and metastasis assays.

Task 2. **Validate differential oncogenic behavior of additional splice variant pairings *in vivo.*** We will test *in vivo* an additional 4-9 splice variant pairings defined in Aim 1, Task 1, Subtasks 1b-1c (e.g. one variant pairing could be the AA-specific and EA-counterpart variants for *FGFR3*), or defined in Aim 1, Task 2, Subtasks 2c-2d.

Year 1 major accomplishments:

i. Full-length cDNA cloning of AA and CA variants of *PIK3CD* and *FGFR3*

We have cloned the full-length cDNA sequences of the AA and CA splice variants for the phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta (*PIK3CD*) and fibroblast growth factor receptor 3 (*FGFR3*) genes (**Figure 1**). Cloning was accomplished by 5'- and 3'-RACE in AA and CA PCa cell lines, and an epitope tag (6×His tag) was attached to the 3'-end of each variant prior to subcloning into the plasmid vector pGEM using standard approaches in our lab. The AA variants of both genes have not been described in the literature nor the UCSC (genome.ucsc.edu) or Ensembl Genome Browsers (www.ensembl.org). The AA *PIK3CD-short* variant is missing exon 20 (encoding a 54 amino acid long segment within the kinase catalytic domain) while the AA *FGFR3-short* variant is missing exon 14 (encoding a 41 amino acid long segment within the tyrosine kinase domain) (**Figure 2**). The CA *PIK3CD* and *FGFR3* versions are long variants, suggesting that kinase activity of the AA variant proteins may differ from the CA variants. To date, the AA and CA cDNA variants of *PIK3CD* and *FGFR3* have been subcloned into the mammalian expression vector pcDNA3, and the *PIK3CD* cDNA variants have been sequence validated and ectopically and stably expressed into PCa cell lines for *in vitro* and *in vivo* investigations completed in Year 1 (see description below).

FIGURE 1. Full-length cDNA sequences of the AA short and CA long variants for *PIK3CD* and *FGFR3*.

>PIK3CD-long CA variant. Sequence shown below is the full length coding sequence of CA variant 3,135 nt, 1,044 aa. PIK3CD-short AA variant is missing exon 20 (which is highlighted in yellow in the CA variant sequence below).

```
atgccccctggggtggactgccccatggaattctggaccaaggaggagaatcagagcgtt
 M  P  P  G  V  D  C  P  M  E  F  W  T  K  E  E  N  Q  S  V
gtggttgacttcctgctgcccacaggggtctacctgaacttccctgtgtcccgcaatgcc
 V  V  D  F  L  L  P  T  G  V  Y  L  N  F  P  V  S  R  N  A
aacctcagcaccatcaagcagctgctgtggcaccgcgcccagtatgagccgctcttccac
 N  L  S  T  I  K  Q  L  L  W  H  R  A  Q  Y  E  P  L  F  H
atgctcagtggccccgaggcctatgtgttcacctgcatcaaccagacagcggagcagcaa
 M  L  S  G  P  E  A  Y  V  F  T  C  I  N  Q  T  A  E  Q  Q
gagctggaggacgagcaacggcgtctgtgtgacgtgcagcccttcctgcccgtcctgcgc
 E  L  E  D  E  Q  R  R  L  C  D  V  Q  P  F  L  P  V  L  R
ctggtggcccgtgagggcgaccgcgtgaagaagctcatcaactcacagatcagcctcctc
 L  V  A  R  E  G  D  R  V  K  K  L  I  N  S  Q  I  S  L  L
atcggcaaaggcctccacgagtttgactccttgtgcgacccagaagtgaacgactttcgc
 I  G  K  G  L  H  E  F  D  S  L  C  D  P  E  V  N  D  F  R
gccaagatgtgccaattctgcgaggaggcggccgcccgccggcagcagctgggctgggag
 A  K  M  C  Q  F  C  E  E  A  A  A  R  R  Q  Q  L  G  W  E
gcctggctgcagtacagtttcccccctgcagctggagccctcggctcaaacctgggggcct
 A  W  L  Q  Y  S  F  P  L  Q  L  E  P  S  A  Q  T  W  G  P
ggtaccctgcggctcccgaaccgggccccttctggtcaacgttaagtttgagggcagcgag
 G  T  L  R  L  P  N  R  A  L  L  V  N  V  K  F  E  G  S  E
gagagcttcaccttccaggtgtccaccaaggacgtgccgctggcgctgatggcctgtgcc
 E  S  F  T  F  Q  V  S  T  K  D  V  P  L  A  L  M  A  C  A
ctgcggaagaaggccacagtgttccggcagccgctggtggagcagccggaagactacacg
 L  R  K  K  A  T  V  F  R  Q  P  L  V  E  Q  P  E  D  Y  T
ctgcaggtgaacggcaggcatgagtacctgtatggcagctacccgctctgccagttccag
 L  Q  V  N  G  R  H  E  Y  L  Y  G  S  Y  P  L  C  Q  F  Q
tacatctgcagctgcctgcacagtgggttgacccctcacctgaccatggtccattcctcc
 Y  I  C  S  C  L  H  S  G  L  T  P  H  L  T  M  V  H  S  S
tccatcctcgccatgcgggatgagcagagcaaccctgccccccaggtccagaaaccgcgt
 S  I  L  A  M  R  D  E  Q  S  N  P  A  P  Q  V  Q  K  P  R
gccaaaccacctcccattcctgcgaagaagccttcctctgtgtccctgtggtccctggag
 A  K  P  P  P  I  P  A  K  K  P  S  S  V  S  L  W  S  L  E
cagccgttccgcatcgagctcatccagggcagcaaagtgaacgccgacgagcggatgaag
 Q  P  F  R  I  E  L  I  Q  G  S  K  V  N  A  D  E  R  M  K
ctggtggtgcaggccgggcttttccacggcaacgagatgctgtgcaagacggtgtccagc
 L  V  V  Q  A  G  L  F  H  G  N  E  M  L  C  K  T  V  S  S
tcggaggtgagcgtgtgctcggagcccgtgtggaagcagcggctggagttcgacatcaac
 S  E  V  S  V  C  S  E  P  V  W  K  Q  R  L  E  F  D  I  N
atctgcgacctgccccgcatggcccgtctctgctttgcgctgtacgccgtgatcgagaaa
 I  C  D  L  P  R  M  A  R  L  C  F  A  L  Y  A  V  I  E  K
gccaagaaggctcgctccaccaagaagaagtccaagaaggcggactgccccattgcctgg
 A  K  K  A  R  S  T  K  K  K  S  K  K  A  D  C  P  I  A  W
gccaacctcatgctgtttgactacaaggaccagcttaagaccggggaacgctgcctctac
 A  N  L  M  L  F  D  Y  K  D  Q  L  K  T  G  E  R  C  L  Y
atgtggccctccgtcccagatgagaagggcgagctgctgaaccccacgggcactgtgcgc
 M  W  P  S  V  P  D  E  K  G  E  L  L  N  P  T  G  T  V  R
agtaaccccaacacggatagcgccgctgccctgctcatctgcctgcccgaggtggccccg
 S  N  P  N  T  D  S  A  A  A  L  L  I  C  L  P  E  V  A  P
caccccgtgtactaccccgccctggagaagatcttggagctggggcgacacagcgagtgt
```

```
 H  P  V  Y  Y  P  A  L  E  K  I  L  E  L  G  R  H  S  E  C
gtgcatgtcaccgaggaggagcagctgcagctgcgggaaatcctggagcggcgggggtct
 V  H  V  T  E  E  E  Q  L  Q  L  R  E  I  L  E  R  R  G  S
ggggagctgtatgagcacgagaaggacctggtgtggaagctgcggcatgaagtccaggag
 G  E  L  Y  E  H  E  K  D  L  V  W  K  L  R  H  E  V  Q  E
cacttcccggaggcgctagcccggctgctgctggtcaccaagtggaacaagcatgaggat
 H  F  P  E  A  L  A  R  L  L  L  V  T  K  W  N  K  H  E  D
gtggcccagatgctctacctgctgtgctcctggccggagctgcccgtcctgagcgccctg
 V  A  Q  M  L  Y  L  L  C  S  W  P  E  L  P  V  L  S  A  L
gagctgctagacttcagcttccccgattgccacgtaggctccttcgccatcaagtcgctg
 E  L  L  D  F  S  F  P  D  C  H  V  G  S  F  A  I  K  S  L
cggaaactgacggacgatgagctgttccagtacctgctgcagctggtgcaggtgctcaag
 R  K  L  T  D  D  E  L  F  Q  Y  L  L  Q  L  V  Q  V  L  K
tacgagtcctacctggactgcgagctgaccaaattcctgctggaccgggccctggccaac
 Y  E  S  Y  L  D  C  E  L  T  K  F  L  L  D  R  A  L  A  N
cgcaagatcggccacttccttttctggcacctccgctccgagatgcacgtgccgtcggtg
 R  K  I  G  H  F  L  F  W  H  L  R  S  E  M  H  V  P  S  V
gccctgcgcttcggcctcatcctggaggcctactgcaggggcagcacccaccacatgaag
 A  L  R  F  G  L  I  L  E  A  Y  C  R  G  S  T  H  H  M  K
gtgctgatgaagcaggggaagcactgagcaaactgaaggccctgaatgacttcgtcaag
 V  L  M  K  Q  G  E  A  L  S  K  L  K  A  L  N  D  F  V  K
ctgagctctcagaagacccccaagccccagaccaaggagctgatgcacttgtgcatgcgg
 L  S  S  Q  K  T  P  K  P  Q  T  K  E  L  M  H  L  C  M  R
caggaggcctacctagaggccctctcccacctgcagtccccactcgaccccagcaccctg
 Q  E  A  Y  L  E  A  L  S  H  L  Q  S  P  L  D  P  S  T  L
ctggctgaagtctgcgtggagcagtgcaccttcatggactccaagatgaagcccctgtgg
 L  A  E  V  C  V  E  Q  C  T  F  M  D  S  K  M  K  P  L  W
atcatgtacagcaacgaggaggcaggcagcggcggcagcgtgggcatcatctttaagaac
 I  M  Y  S  N  E  E  A  G  S  G  G  S  V  G  I  I  F  K  N
ggggatgacctccggcaggacatgctgaccctgcagatgatccagctcatggacgtcctg
 G  D  D  L  R  Q  D  M  L  T  L  Q  M  I  Q  L  M  D  V  L
tggaagcaggaggggctggacctgaggatgaccccctatggctgcctccccaccggggac
 W  K  Q  E  G  L  D  L  R  M  T  P  Y  G  C  L  P  T  G  D
cgcacaggcctcattgaggtggtactccgttcagacaccatcgccaacatccaactcaac
 R  T  G  L  I  E  V  V  L  R  S  D  T  I  A  N  I  Q  L  N
aagagcaacatggcagccacagccgccttcaacaaggatgccctgctcaactggctgaag
 K  S  N  M  A  A  T  A  A  F  N  K  D  A  L  L  N  W  L  K
tccaagaacccgggggaggccctggatcgagccattgaggagttcaccctctcctgtgct
 S  K  N  P  G  E  A  L  D  R  A  I  E  E  F  T  L  S  C  A
ggctattgtgtggccacatatgtgctgggcattggcgatcggcacagcgacaacatcatg
 G  Y  C  V  A  T  Y  V  L  G  I  G  D  R  H  S  D  N  I  M
atccgagagagtgggcagctgttccacattgattttggccactttctggggaatttcaag
 I  R  E  S  G  Q  L  F  H  I  D  F  G  H  F  L  G  N  F  K
accaagtttggaatcaaccgcgagcgtgtcccattcatcctcacctacgactttgtccat
 T  K  F  G  I  N  R  E  R  V  P  F  I  L  T  Y  D  F  V  H
gtgattcagcaggggaagactaataatagtgagaaatttgaacggttccggggctactgt
 V  I  Q  Q  G  K  T  N  N  S  E  K  F  E  R  F  R  G  Y  C
gaaagggcctacaccatcctgcggcgccacgggcttctcttcctccacctctttgccctg
 E  R  A  Y  T  I  L  R  R  H  G  L  L  F  L  H  L  F  A  L
atgcgggcggcaggcctgcctgagctcagctgctccaaagacatccagtatctcaaggac
 M  R  A  A  G  L  P  E  L  S  C  S  K  D  I  Q  Y  L  K  D
tccctggcactggggaaaacagaggaggaggcactgaagcacttccgagtgaagtttaac
 S  L  A  L  G  K  T  E  E  E  A  L  K  H  F  R  V  K  F  N
gaagccctccgtgagagctggaaaaccaaagtgaactggctggcccacaacgtgtccaaa
 E  A  L  R  E  S  W  K  T  K  V  N  W  L  A  H  N  V  S  K
gacaacaggcagtag
 D  N  R  Q  -
```

>FGFR3-long CA variant. Sequence shown below is the full length coding sequence of CA variant 2,427 nt, 809 aa. FGFR3-short AA variant is missing exon 14 (which is highlighted in yellow in the CA variant sequence below).

```
atgggcgcccctgcctgcgccctcgcgctctgcgtggccgtggccatcgtggccggcgcc
 M  G  A  P  A  C  A  L  A  L  C  V  A  V  A  I  V  A  G  A
tcctcggagtccttggggacggagcagcgcgtcgtggggcgagcggcagaagtcccgggc
 S  S  E  S  L  G  T  E  Q  R  V  V  G  R  A  A  E  V  P  G
ccagagcccggccagcaggagcagttggtcttcggcagcggggatgctgtggagctgagc
 P  E  P  G  Q  Q  E  Q  L  V  F  G  S  G  D  A  V  E  L  S
tgtcccccgcccgggggtggtcccatggggcccactgtctgggtcaaggatggcacaggg
 C  P  P  P  G  G  G  P  M  G  P  T  V  W  V  K  D  G  T  G
ctggtgccctcggagcgtgtcctggtggggccccagcggctgcaggtgctgaatgcctcc
 L  V  P  S  E  R  V  L  V  G  P  Q  R  L  Q  V  L  N  A  S
cacgaggactccggggcctacagctgccggcagcggctcacgcagcgcgtactgtgccac
 H  E  D  S  G  A  Y  S  C  R  Q  R  L  T  Q  R  V  L  C  H
ttcagtgtgcgggtgacagacgctccatcctcgggagatgacgaagacggggaggacgag
 F  S  V  R  V  T  D  A  P  S  S  G  D  D  E  D  G  E  D  E
gctgaggacacaggtgtggacacaggggccccttactggacacggcccgagcggatggac
 A  E  D  T  G  V  D  T  G  A  P  Y  W  T  R  P  E  R  M  D
aagaagctgctggccgtgccggccgccaacaccgtccgcttccgctgcccagccgctggc
 K  K  L  L  A  V  P  A  A  N  T  V  R  F  R  C  P  A  A  G
aaccccactccctccatctcctggctgaagaacggcagggagttccgcggcgagcaccgc
 N  P  T  P  S  I  S  W  L  K  N  G  R  E  F  R  G  E  H  R
attggaggcatcaagctgcggcatcagcagtggagcctggtcatggaaagcgtggtgccc
 I  G  G  I  K  L  R  H  Q  Q  W  S  L  V  M  E  S  V  V  P
tcggaccgcggcaactacacctgcgtcgtggagaacaagtttggcagcatccggcagacg
 S  D  R  G  N  Y  T  C  V  V  E  N  K  F  G  S  I  R  Q  T
tacacgctggacgtgctggagcgctccccgcaccggcccatcctgcaggcggggctgccg
 Y  T  L  D  V  L  E  R  S  P  H  R  P  I  L  Q  A  G  L  P
gccaaccagacggcggtgctgggcagcgacgtggagttccactgcaaggtgtacagtgac
 A  N  Q  T  A  V  L  G  S  D  V  E  F  H  C  K  V  Y  S  D
gcacagccccacatccagtggctcaagcacgtggaggtgaatggcagcaaggtgggcccg
 A  Q  P  H  I  Q  W  L  K  H  V  E  V  N  G  S  K  V  G  P
gacggcacaccctacgttaccgtgctcaagtcctggatcagtgagagtgtggaggccgac
 D  G  T  P  Y  V  T  V  L  K  S  W  I  S  E  S  V  E  A  D
gtgcgcctccgcctggccaatgtgtcggagcgggacgggggcgagtacctctgtcgagcc
 V  R  L  R  L  A  N  V  S  E  R  D  G  G  E  Y  L  C  R  A
accaatttcataggcgtggccgagaaggccttttggctgagcgttcacgggccccgagca
 T  N  F  I  G  V  A  E  K  A  F  W  L  S  V  H  G  P  R  A
gccgaggaggagctggtggaggctgacgaggcgggcagtgtgtatgcaggcatcctcagc
 A  E  E  E  L  V  E  A  D  E  A  G  S  V  Y  A  G  I  L  S
tacggggtgggcttcttcctgttcatcctggtggtggcggctgtgacgctctgccgcctg
 Y  G  V  G  F  F  L  F  I  L  V  V  A  A  V  T  L  C  R  L
cgcagccccccccaagaaaggcctgggctccccccaccgtgcacaagatctcccgcttcccg
 R  S  P  P  K  K  G  L  G  S  P  T  V  H  K  I  S  R  F  P
ctcaagcgacaggtgtccctggagtccaacgcgtccatgagctccaacacaccactggtg
 L  K  R  Q  V  S  L  E  S  N  A  S  M  S  S  N  T  P  L  V
cgcatcgcaaggctgtcctcaggggagggccccacgctggccaatgtctccgagctcgag
 R  I  A  R  L  S  S  G  E  G  P  T  L  A  N  V  S  E  L  E
ctgcctgccgaccccaaatgggagctgtctcgggcccggctgaccctgggcaagcccctt
 L  P  A  D  P  K  W  E  L  S  R  A  R  L  T  L  G  K  P  L
ggggagggctgcttcggccaggtggtcatggcggaggccatcggcattgacaaggaccgg
 G  E  G  C  F  G  Q  V  V  M  A  E  A  I  G  I  D  K  D  R
gccgccaagcctgtcaccgtagccgtgaagatgctgaaagacgatgccactgacaaggac
```

```
 A  A  K  P  V  T  V  A  V  K  M  L  K  D  D  A  T  D  K  D
ctgtcggacctggtgtctgagatggagatgatgaagatgatcgggaaacacaaaaacatc
 L  S  D  L  V  S  E  M  E  M  M  K  M  I  G  K  H  K  N  I
atcaacctgctgggcgcctgcacgcagggcgggcccctgtacgtgctggtggagtacgcg
 I  N  L  L  G  A  C  T  Q  G  G  P  L  Y  V  L  V  E  Y  A
gccaagggtaacctgcgggagtttctgcgggcgcggcggccccgggcctggactactcc
 A  K  G  N  L  R  E  F  L  R  A  R  R  P  P  G  L  D  Y  S
ttcgacacctgcaagccgcccgaggagcagctcaccttcaaggacctggtgtcctgtgcc
 F  D  T  C  K  P  P  E  E  Q  L  T  F  K  D  L  V  S  C  A
taccaggtggcccggggcatggagtacttggcctcccagaagtgcatccacagggacctg
 Y  Q  V  A  R  G  M  E  Y  L  A  S  Q  K  C  I  H  R  D  L
gctgcccgcaatgtgctggtgaccgaggacaacgtgatgaagatcgcagacttcgggctg
 A  A  R  N  V  L  V  T  E  D  N  V  M  K  I  A  D  F  G  L
gcccgggacgtgcacaacctcgactactacaagaagacgaccaacggccggctgcccgtg
 A  R  D  V  H  N  L  D  Y  Y  K  K  T  T  N  G  R  L  P  V
aagtggatggcgcctgaggccttgtttgaccgagtctacactcaccagagtgacgtctgg
 K  W  M  A  P  E  A  L  F  D  R  V  Y  T  H  Q  S  D  V  W
tcctttggggtcctgctctgggagatcttcacgctgggggctccccgtaccccggcatc
 S  F  G  V  L  L  W  E  I  F  T  L  G  G  S  P  Y  P  G  I
cctgtggaggagctcttcaagctgctgaaggagggccaccgcatggacaagcccgccaac
 P  V  E  E  L  F  K  L  L  K  E  G  H  R  M  D  K  P  A  N
tgcacacacgacctgtacatgatcatgcgggagtgctggcatgccgcgccctcccagagg
 C  T  H  D  L  Y  M  I  M  R  E  C  W  H  A  A  P  S  Q  R
cccaccttcaagcagctggtggaggacctggaccgtgtccttaccgtgacgtccaccgac
 P  T  F  K  Q  L  V  E  D  L  D  R  V  L  T  V  T  S  T  D
gagtacctggacctgtcggcgcctttcgagcagtactccccgggtggccaggacacccccc
 E  Y  L  D  L  S  A  P  F  E  Q  Y  S  P  G  G  Q  D  T  P
agctccagctcctcaggggacgactccgtgtttgcccacgacctgctgcccccggcccca
 S  S  S  S  S  G  D  D  S  V  F  A  H  D  L  L  P  P  A  P
cccagcagtgggggctcgcggacgtga
 P  S  S  G  G  S  R  T  -
```

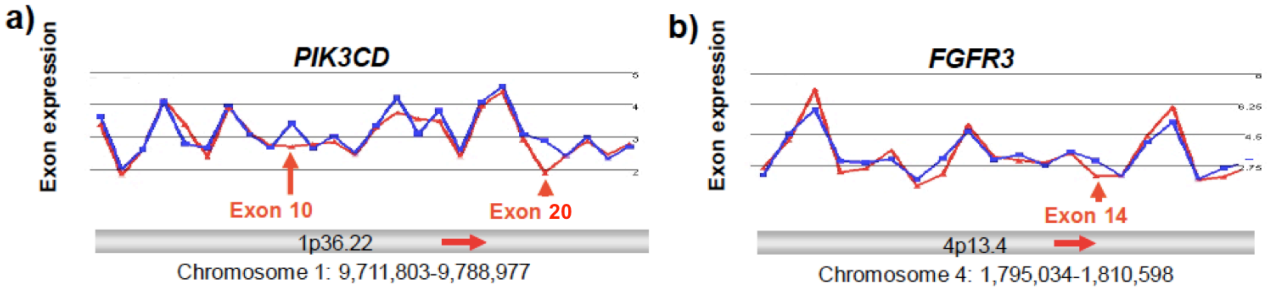

FIGURE 2. Affymetrix exon array profiling. (a) PIK3CD-long variant expressed in CA PCa cell lines has both exons 10 and 20. PIK3CD-short variant expressed in AA PCa cell lines is missing exons 10 and/or 20. **(b)** FGFR3-long variant expressed in CA PCa cell lines includes exon 14. FGFR3-short variant expressed in AA PCa cell lines is missing exon 14.

ii. **Validation that the AA short variant of *PIK3CD* is phenotypically more aggressive than the CA long variant based on *in vitro* assays.** Assays performed included genetic manipulation of AA and CA variant expression in PCa cell lines, proliferation assays, and invasion assay in Matrigel.

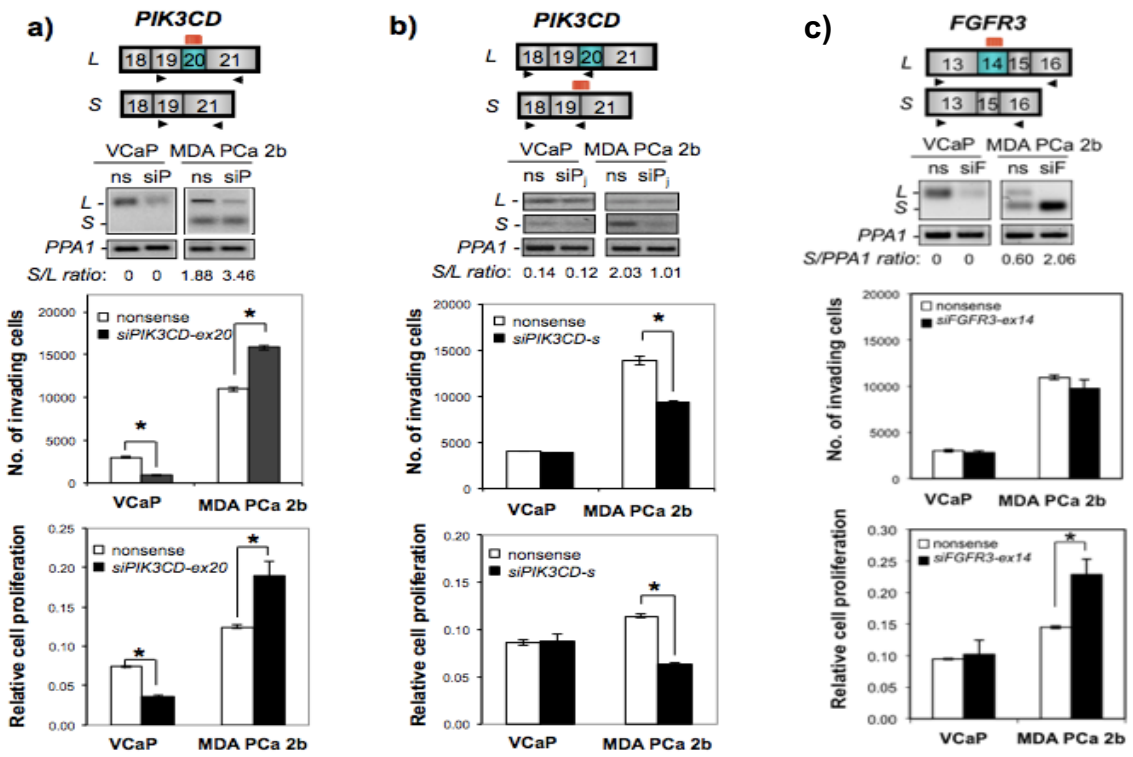

FIGURE 3. Functional consequences of targeted knockdown of splice variants of PIK3CD in AA and CA PCa cell lines. **(a)** Effects of siRNA-mediated knockdown of PIK3CD long splice variant (L) found in CA PCa cell line VCaP and AA PCa cell line MDA PCa 2b. The L variant is found expressed in both VCaP and MDA PCa 2b cells. **(b)** Effects of siRNA-mediated knockdown of PIK3CD short splice variant (S) found in AA PCa cell line MDA PCa 2b. The EA PCa cell line VCaP expresses very little to no S variant. Knockdown of the L variant was accomplished with an exon 20-specific siRNA, while knockdown of the S variant was accomplished by an siRNA spanning exons 19 and 21. **(c)** Effects of siRNA-mediated knockdown of FGFR3 long splice variant (L) found in CA PCa cell line VCaP and AA PCa cell line MDA PCa 2b. The L variants are found expressed in both VCaP and MDA PCa 2b cells. Data are the mean ± SE of 4 independent experiments.

We demonstrate that the AA short splice variant (S variant) for *PIK3CD* encodes a more aggressive version of the gene (i.e. leading to greater proliferation and invasion of cancer cells) compared to the CA long (L) variant counterpart. SiRNA-mediated knockdown of the L variant in CA PCa cell line VCaP (**Figure 3A**, middle panel qRT-PCR, siP lane) leads to a decrease in Matrigel invasion and a decrease in proliferation as assessed by BrdU incorporation (**Figure 3A**, bottom panels). By comparison, the AA PCa cell line MDA PCa 2b expresses both L and S variants, and knockdown of the L variant leads to predominant expression of the S variant and a corresponding increase in Matrigel invasion and increase in proliferation (**Figure 3A**, middle and bottom panels). Next, we investigated S variant knockdown. VCaP cells express little to no S variant; hence, targeted siRNA-mediated knockdown of this variant led to no change in Matrigel invasion and proliferation (**Figure 3B**). In contrast, targeted knockdown of the S variant in MDA PCa 2b cells leads to decreased Matrigel invasion and decreased proliferation (knockdown of S variant leads to predominant expression of L variant) (**Figure 3B**). These data indicate that the overall S to L ratio in MDA PCa 2b cells dictates the oncogenic profile of this AA PCa cell line. Namely, knocking down the L variant in MDA PCa 2b cells increases the S/L ratio, leading to a higher proportion of the aggressive S variant and consequently increased invasiveness and

proliferation of the cell line. In contrast, knocking down the S variant in MDA PCa 2b cells decreases the S/L ratio, leading to a higher proportion of the less aggressive L variant and consequently decreased invasiveness and proliferation of the cell line. An analogous increased proliferative behavior was obtained when the AA-'specific' variant (found exclusively or nearly exclusively in AA cell lines) to CA-counterpart variant (found in both EA and AA cell lines) ratio was increased for FGFR3 (**Figure 3C**).

The more aggressive invasive behavior observed in the AA PCa cell line MDA PCa 2b upon increasing the S/L variant ratio was associated with an augmented activation of the PI3K/AKT pathway. This was evidenced by the increased phosphorylation of AKT at amino acids Thr308 and Ser473, mTOR and ribosomal protein S6 (S6) (**Figure 4A**).

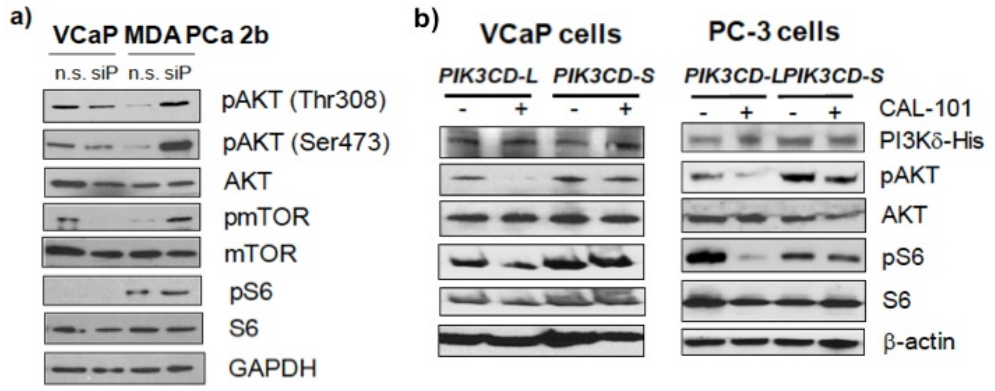

FIGURE 4. SiRNA-mediated knockdown or pharmacological inhibition reduces activity of the PI3K/AKT signaling in PIK3CD-L- but not in PIK3CD-S-expressing cells. (a) CA VCaP and AA MDA PCa 2b cells were transfected with nonsense siRNA (n.s.) or siRNA against exon 20 of PIK3CD-L (siP). Western blot analysis revealed that siP knockdown caused a decrease in the phosphorylation status (i.e. activation) of AKT (pAKT), mTOR (pmTOR) and S6 (pS6) in VCaP cells, while siP transfection increased phosphorylation of these signaling proteins in MDA PCa 2b. **(b)** Phosphorylation status of AKT, mTOR and S6 proteins are inhibited by CAL-101 in VCaP and PC-3 cells over-expressing the PIK3CD-L variant but not the PIK3CD-S variant. Data are the mean ± SE of 4 independent experiments.

iii. **Demonstration in our *in vitro* assays that the AA variant of PIK3CD is resistant to inhibition by CAL-101, a small molecule inhibitor designed specifically against PIK3CD.** In contrast, the CA variant of PIK3CD is sensitive to inhibition by CAL-101

We subsequently stably over-expressed the AA PIK3CD-short (PIK3CD-S) and CA PIK3CD-long variants (PIK3CD-L) variants individually in the CA PCa cell lines PC-3 and VCaP. These stably transfected cell lines were tested *in vitro* for sensitivity to CAL-101 treatment (**Figure 4B**). CAL-101 is a PIK3CD inhibitor in clinical trials for various cancers. PCa cell lines over-expressing the CA PIK3CD-L variant exhibited a decrease in the activity of the PI3K/AKT pathway following CAL-101 treatment, as seen by a loss of AKT, mTOR and S6 phosphorylation. Remarkably, the same CA PCa cell lines stably over-expressing equivalent levels of the AA PIK3CD-S variant were completely resistant to CAL-101. In other words, there was no significant change in AKT, mTOR and S6 phosphorylation levels before and after CAL-101 treatment.

In cell proliferation assays, we demonstrate that BrdU labeling in VCaP and PC-3 cells over-expressing the CA PIK3CD-L variant was inhibited by CAL-101 in a dose-dependent manner, while proliferation of VCaP and PC-3 cells over-expressing the AA PIK3CD-L variant were resistant to CAL-101 (**Figure 5**).

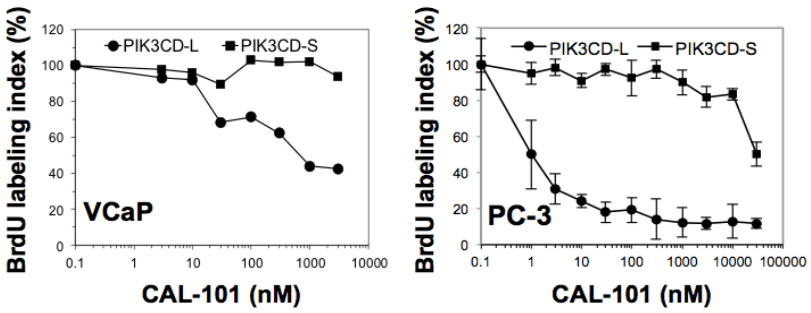

FIGURE 5. CAL-101 treatment inhibits proliferation of PIK3CD-L- but not in PIK3CD-S-expressing VCaP (left panel) and PC-3 cells (right panel). Data are the mean ± SE of 4 independent experiments.

Year 2 major accomplishments:

i. **Manuscript submitted to *Nature Communications* detailing our findings on the oncogenic behavior of the AA-specific/enriched *PIK3CD* short variant (*PIK3CD-S*) compared to the EA *PIK3CD* long variant (*PIK3CD-L*).**

We completed our *in vitro* and *in vivo* studies comparing the S and L variants of PI3KCD at the end of Year 2 and submitted the manuscript to *Nature Communications*. Findings in Year 2 can be summarized as follows:

Cell-free system: PI3KCD-S activity is resistant to CAL-101, while EA variant PI3KCD-L is sensitive. We have recently purified His-tagged PIK3CD-S and -L using a HisPur Ni-NTA column approach (**Figure 6A**), and subjected the purified proteins to a cell-free *in vitro* colorimetric-based PI3K activity/inhibitor assay (**Figure 6B**). Interestingly, baseline activities of the S and L isoforms were not significantly different. Notwithstanding, our results clearly demonstrate that the S isoform was resistant to inhibition by CAL-101 and wortmannin, whereas L isoform activity was completely inhibited by these small molecule inhibitors.

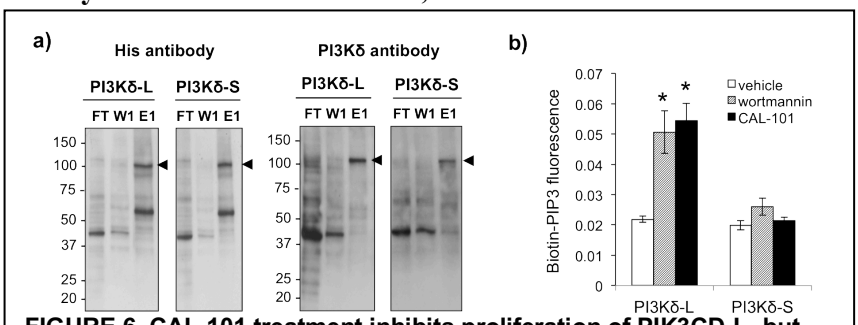

FIGURE 6. CAL-101 treatment inhibits proliferation of PIK3CD-L- but not in PIK3CD-S-expressing VCaP (left panel) and PC-3 cells (right panel). Data are the mean ± SE of 4-6 independent experiments.

Xenograft of PC-3 cells over-expressing PIK3CD-S is resistant, while PC-3 over-expressing PIK3CD-L is sensitive, to the anti-proliferative and anti-metastatic effects of CAL-101. We demonstrate in a xenograft mouse model that proliferative growth and metastasis of PC-3

cells over-expressing the AA-specific/enriched PIK3CD-S variant were resistant to the inhibitory

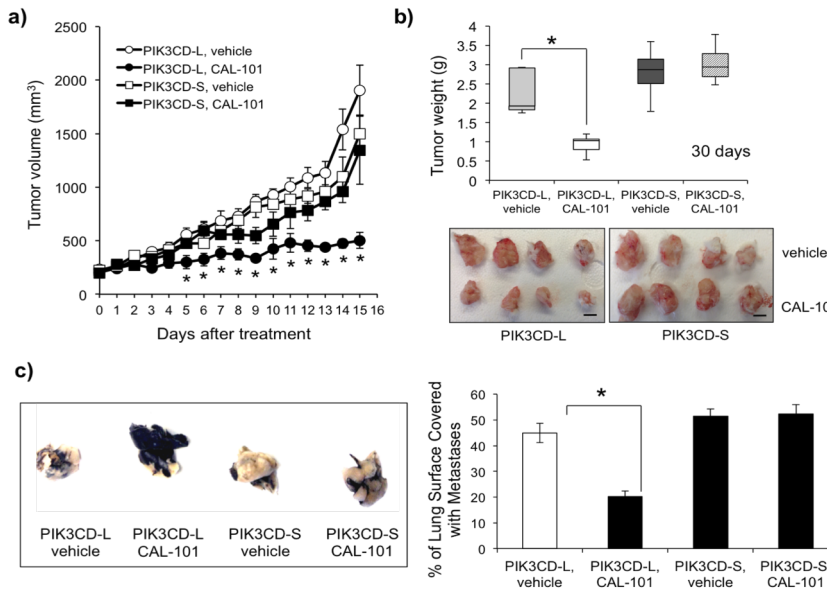

FIGURE 7. Differential response of *PI3KCD-L* and *PI3KCD-S* isoforms to CAL-101 treatment *in vivo*. (**a**) PC-3 cells stably expressing *PIK3CD-L* exhibited reduced tumorigenesis in NOD-SCID mice upon CAL-101 treatment. In contrast, PC-3 cells stably expressing *PIK3CD-S* exhibited resistance to CAL-101 inhibition of tumorigenesis in NOD-SCID mice. Two million PC-3 cells expressing *PIK3CD-L* or *PIK3CD-S* were injected subcutaneously into the left flanks of each NOD-SCID mice and allowed to develop tumors under vehicle or CAL-101 treatment. The data represent the mean tumor size ± SEM, with * $P < 0.05$ (ANOVA, n=10 independent mice for each treatment group at each time point). (**b**) Tumor weights and gross morphology of the tumor xenografts from (**a**). The box plots represent mean tumor weights after 15-days vehicle or CAL-101 treatment (ANOVA, n=5 independent mice for each treatment group at each time point). (**c**) Intravenous tail vein with 10^5 PC-3 cells ectopically expressing *PIK3CD*-S or *PIK3CD*-L for lung metastasis assay. Animals treated with daily IP injections of vehicle (veh), PC-3 cells expressing either variant grew and metastasized robustly. Animals treated with daily IP injections of CAL-101 (CAL; 25 mg/kg), cells expressing S variant still exhibited robust growth and metastasis, while L expressing cells exhibited significant inhibition of growth and metastasis. (*P<0.05, significantly different from vehicle treated cells expressing L; ANOVA with post-hoc test). n=10 for each exp. group. Representative xenograft growths and metastases (white masses).

effects of CAL-101. By comparison, PC-3 cells over-expressing the EA PIK3CD-L variant were particularly sensitive to CAL-101 treatment (**Figure 7**).

Molecular modeling of PIK3CD-S and -L protein variants. Molecular modeling demonstrates that CAL-101 is unable to dock onto the ATP binding pocket of PIK3CD-S due to the absence of key amino acids Glu826 and Val828 (these amino acids are encoded by exon 20 that is missing in PIK3CD-S), whereas CAL-101 efficiently docks and consequently would inhibit kinase activity of PIK3CD-L (**Figure 8**). These findings provide an explanation for the resistance of PIK3CD-S to small molecule inhibitors. We anticipate that future studies will be aimed at developing inhibitors specific to PIK3CD-S.

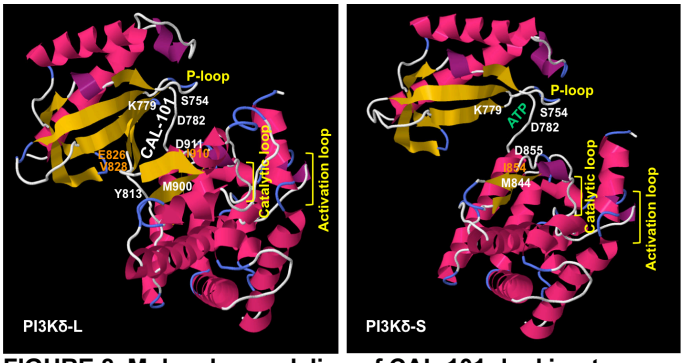

FIGURE 8. Molecular modeling of CAL-101 docking to PIK3CD-L but not to PIK3CD-S.

Years 3 major accomplishments:

i. **Manuscript detailing our findings on the oncogenic behavior of the AA-specific/enriched *PIK3CD* short variant (*PIK3CD-S*) compared to the EA *PIK3CD* long variant (*PIK3CD-L*) had been positively reviewed by *Nature Communications***

Assessment of our *Nature Communications* manuscript by the reviewers had been positively reviewed at the end of Year 2, albeit they requested a number of additional experiments that were subsequently completed during Year 3. Our manuscript was resubmitted at the end of Year 3. The additional findings can be summarized as follows:

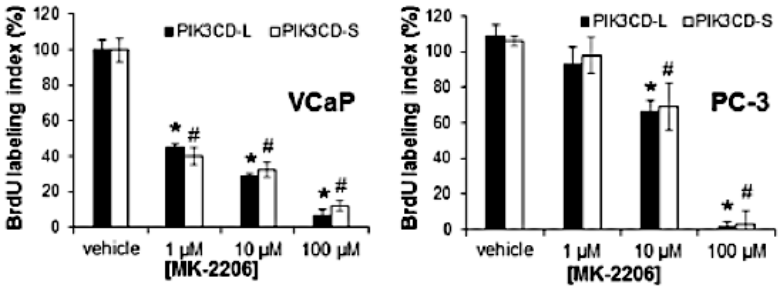

FIGURE 9. PI3Kδ-S oncogenic signaling, which is resistant to small molecule inhibitor CAL-101, is sensitive to downstream inhibition of AKT. Treatment with vehicle (saline) or selective AKT small molecule inhibitor MK-2206 (24 hrs). Proliferation was assessed using a BrdU labeling assay. Data presented as the mean ± SEM of n=4-6 independent experiments for each treatment group. * or #Significantly different from corresponding vehicle control, $P < 0.05$ by ANOVA with Dunnett's post-hoc test.

AA-specific/enriched variant PI3KCD-S imparts oncogenic behavior that is resistant to small molecule inhibitor CAL-101, which can be overcome by AKT inhibition. The reviewers had requested that we assess whether or not cell lines over-expressing the *PIK3CD-S* or *PIK3CD-L* variant exhibited differential sensitivity to inhibitors of downstream signaling components. Interestingly, *PIK3CD-S*- and *PIK3CD-L*-over-expressing cell lines were equally to selective inhibition of AKT with MK-2206 (**Figure 9**). These results demonstrate that PI3Kδ-S-stimulated proliferation is resistant to CAL-101 inhibition in sharp contrast to PI3Kδ-L; while inhibition of AKT, which is downstream of PI3Kδ-S, effectively blocked proliferation.

ii. **Defining mechanism of PI3KCδ-S oncogenecity and resistance to CAL-101 inhibition**

Employing cell-free system to define: a) mechanism of increased oncogenecity of PI3KCδ-S over PI3KCδ-L, and b) the mechanism of PI3KCδ-S resistance versus PI3KCδ-L sensitivity to CAL-101 inhibition. A major question that the reviewers of *Nature Communications* required us to address was the mechanism of enhanced PI3KCδ-S oncogenecity and resistance (**Figure 10**; note that portions of this figure include portions of experiments completed in Years 1 and 2). The result of the exclusion of exon 20 (168 bp) in the *PIK3CD-S* variant is an in-frame deletion of 56 amino acids (residues 810-865) in the catalytic domain of the PI3Kδ-S isoform (**Figure 10a**). To gain further insight into the functional differences between PI3Kδ isoforms, the interaction of PI3Kδ-L and -S with regulatory subunit p85α was investigated. Whole cell lysates from PC-3 cells over-expressing either His-tagged PI3Kδ-S or PI3Kδ-L were subjected to western analysis, demonstrating that each cell line expressed equivalent levels of their respective PI3Kδ isoform as well as equal p85α expression (**Figure 10b**; left panel). Interestingly, co-IP of the PI3Kδ/p85α complex from whole cell lysates using an anti-His antibody demonstrated that p85α bound with 3-4-fold greater proficiency to

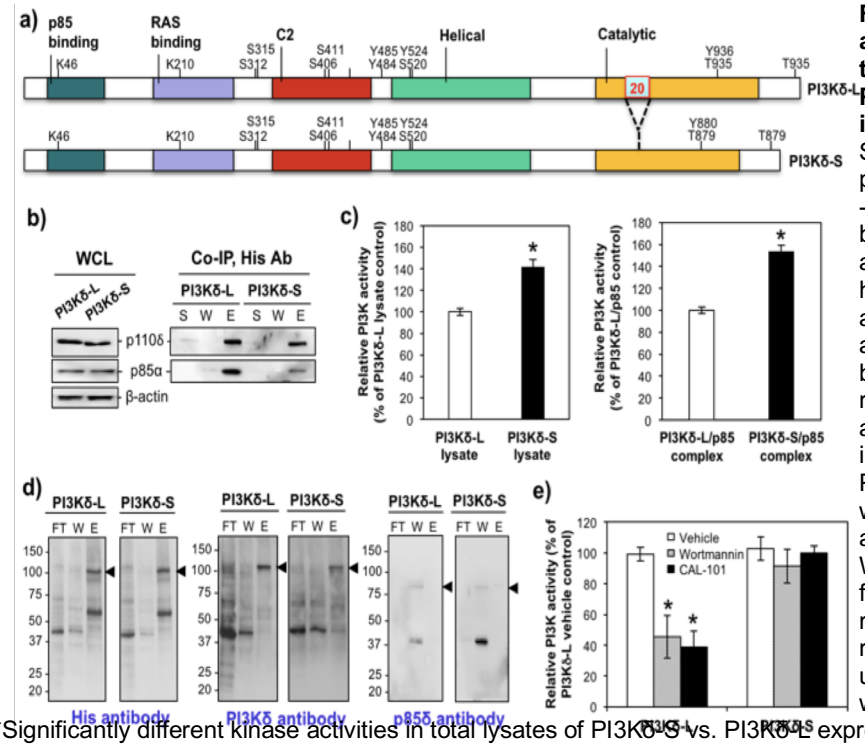

FIGURE 10. Cell-free kinase assay of PI3Kδ isoforms and tertiary structure modeling of PIK3δ-small molecule inhibitor interactions. (a) Schematic representation of protein domains of PI3Kδ-L and -S isoforms. Adaptor (p85)-binding, RAS binding, C2, helix and catalytic domains are highlighted. Phosphorylation and ubiquitination sites (S, T, Y and K) and the region encoded by exon 20 (56 amino acids) residing in the catalytic domain are indicated. (b) Co-immunoprecipitation (Co-IP) of PI3Kδ/p85 complex followed by western blot, (c) and PI3K activity assays. S, supernatant; W, wash fraction; E, eluted fraction. Blots are representative of 3-5 similar results. Anti-His antibody was used independent experiments with similar results. *Significantly different kinase activities in total lysates of PI3Kδ-S vs. PI3Kδ-L expressing cells, or purified PI3Kδ-S/p85 vs. PI3Kδ-L/p85 complexes. $P < 0.05$ using Student t-test. Data presented as mean ± SEM of n = 4 for each treatment group. (d) Purification of His-tagged PI3Kδ-L and -S isoforms. Western blot analysis of Ni-NTA resin-purified PI3Kδ isoforms from transfected PC-3 cells using His and PI3Kδ antibodies. FT, flow-through; W, wash fraction; E, eluted fraction. Blots are representative of 4 independent experiments with similar results. Closed arrowheads indicate PI3Kδ isoforms. (e) Cell-free kinase assay of L and S isoforms of PI3Kδ in the presence of vehicle (PBS), 100nM wortmannin or 100nM CAL-101. *Significantly different from vehicle control-treated PI3Kδ-L isoform. $P < 0.05$ by ANOVA with Dunnett's post-hoc test. Data presented as mean ± SEM of n = 4-5 for each treatment group.

PI3Kδ-L compared to p85α binding to PI3Kδ-S (**Figure 10b**; right panel, column E). Binding proficiency was inversely correlated with PI3Kδ isoform kinase activity (**Figure 10c**; right panel).

Next, His-tagged PI3Kδ isoforms were purified from the lysates of PC-3 cells overexpressing either PI3Kδ-S or PI3Kδ-L using Ni-NTA resin columns. As shown in **Figure 10d** (left and middle panels), purification of the PI3Kδ-S and -L isoforms was verified by western blotting using anti-His or anti-PI3Kδ antibodies. In addition, the Ni-NTA resin column purification approach resulted in the isolation of PI3Kδ isoforms that were no longer bound to p85α (**Figure 10d**; far right panel, column E). The purified PI3Kδ isoforms (minus p85α)

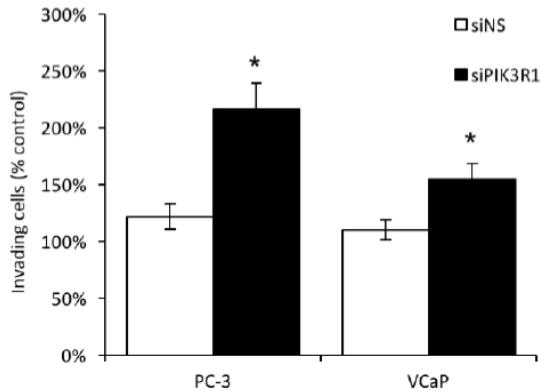

FIGURE 11. Knockdown of p85α increases invasive activity of EA PCa cell lines. Invasion was assessed using Matrigel assay following siRNA-mediated knockdown of p85**a** (PIK3R1). SiRNA-SMART pool for targeting PIK3R1 was purchased from Dharmacon (Lafayette, CO). Data presented as the mean ± SEM of n=3-4 independent experiments for each treatment group. *P < 0.05 by ANOVA with Dunnett's post-hoc test. siNS, Nonsense siRNA control.

were incubated with vehicle, non-selective PI3K inhibitor wortmannin (100 nM) or PI3Kδ-specific inhibitor CAL-101 (100 nM), and subjected to a PI3K activity assay. In the absence of bound p85α, kinase activity of PI3Kδ-L was equivalent to PI3Kδ-S (**Figure 10e**, compare vehicle treatments). In agreement, siRNA-mediated knockdown of p85α in wild-type EA PCa cell lines VCaP and PC-3 was associated with an increase in invasive activity (**Figure 11**). Remarkably, wortmannin and CAL-101 significantly inhibited the activity of the PI3Kδ-L isoform, but not the PI3Kδ-S isoform (**Figure 10e**). These results demonstrate that PI3Kδ-S maintains kinase activity even in the presence of small molecule inhibitors, supporting our earlier *in-vitro* and *in-vivo* results presented in our Year 2 progress report.

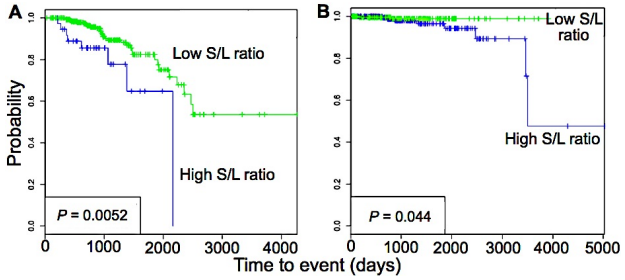

FIGURE 10. PIK3CD-S (A) and FGFR3-S (B) are associated with decreased survival among PCa patients. RNA-Seq and disease free survival data for PCa (n = 494 patients) was obtained from The Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov). P-values for survival curves determined by the log-rank test.

iii. We have determined that high *PI3KCD-S* and *FGFR3-S* expression is associated with poorer patient prognosis

We have performed survival analysis of The Cancer Genome Atlas (TCGA) RNA-Seq data, demonstrating that either high expression of *PIK3CD-S* or *FGFR3-S* associates with significantly poorer survival in PCa patients (**Figure 10**), demonstrating further the potential relevance of these two short variants in PCa.

<u>Year 4 major accomplishments:</u>

i. Manuscript detailing our findings on the oncogenic behavior of the AA-specific/enriched *PIK3CD* short variant (*PIK3CD-S*) compared to the EA *PIK3CD* long variant (*PIK3CD-L*) was published in *Nature Communications.*

ii. *FGFR3-S* is resistant to tyrosine kinase inhibitor dovitinib, while FGFR3-L is sensitive.

EA PCa cell lines PC-3 and LNCaP were transfected for ectopic over-expression of FGFR3-S and FGFR3-L, and assayed for proliferation (**Figure 11**) and Matrigel invasion (**Figure 12**). For proliferation, PCa cell lines over-expressing FGFR3-S, but not FGFR3-L, exhibited higher baseline proliferation compared to PCa cell lines transfected with lacZ control gene (**Figure 11, right panel**). Interestingly, proliferation of S isoform-over-expressing cells demonstrated resistance to TKI dovitinib compared to L isoform-over-expressing cells (**Figure 11, middle and right panels**).

For invasion, PCa cell lines over-expressing either FGFR3-S or FGFR3-L exhibited higher baseline invasion compared to PCa cell lines transfected with lacZ control gene (**Figure 12, right panel**). Invasion of S isoform-over-expressing cells exhibited slight albeit significant resistance to TKI dovitinib compared to L isoform-over-expressing cells (**Figure 12, middle and right panels**).

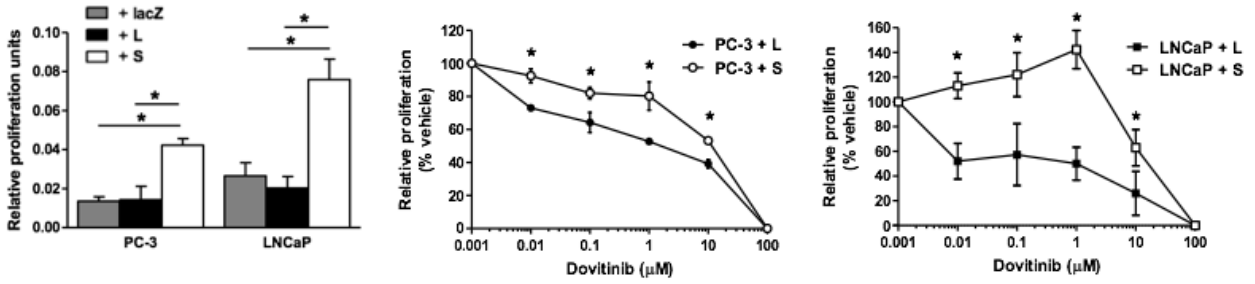

Fig. 2. Proliferation of EA cell lines overexpressing FGFR3-S and FGFR3-L when treated with increasing concentrations of dovitinib. n=3-5, t-test *p<0.05 (Left panel). One-way ANOVA with Tukey's post-hoc, *p<0.05 (Middle and Left panels).

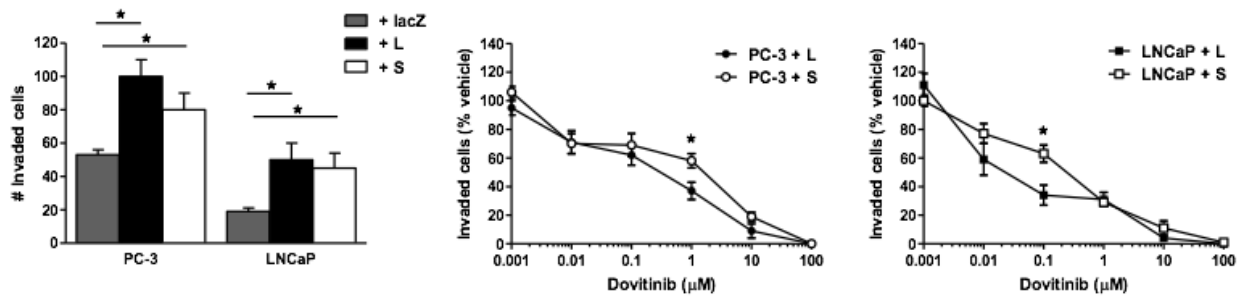

Fig. 3. Invasion of EA cell lines overexpressing FGFR3-S and FGFR3-L when treated with increasing concentrations of dovitinib. n=3-5, t-test *p<0.05 (Left panel). One-way ANOVA with Tukey's post-hoc, *p<0.05 (Middle and Left panels).

Opportunities for training and professional development:

2016 AACR Health Disparities Presentation. This proposal has provided hands-on training for PhD graduate student Jacqueline Olender. The PI is serving as Ms. Olender's mentor and she has participated in both the *in vitro* and *in vivo* work described herein. This work is part of Ms. Olender's PhD dissertation research project. An NIH/AACR Scholar-in-Training Award was awarded to Ms. Olender to present her thesis work entitled, "Alternative splicing of FGFR3 as a mechanism for prostate cancer health disparities" at the 9th AACR Conference on The Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved.

2017 AACR Health Disparities Presentation. An NIH/AACR Scholar-in-Training Award was awarded to Ms. Olender to present her thesis work entitled, "Novel FGFR3 splice variant increases oncogenic phenotype in African American prostate cancer" at the 10th AACR Conference on The Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved.

Dissemination of results and outreach to communities of interest:

Mentoring. During the final year of this grant, my lab mentored Fulbright Fellow Chaimae Samtal (PhD candidate) from the University of Sidi Mohamed Ben, Morocco. Ms. Samtal is a PhD candidate studying the genomics of prostate cancer in the Moroccan population.

Goals for next reporting period:

N/A

4. IMPACT

Impact on the development of the principal discipline(s) of the project:

Principal discipline -- Understanding prostate cancer biology and disparities. Taken together, our *in vitro* and *in vivo* research on AA-specific/enriched *PIK3CD-S* and *FGFR3-S* variants provide strong evidence that differential splicing may play a critical role in PCa health disparities. Our future goal is to identify the underlying molecular mechanism(s) responsible for the differential splicing events observed in AA versus EA PCa specimens.

Impact on other disciplines:

Other disciplines -- Our previous work demonstrated that the short variant encodes a more aggressive oncogenic signaling protein isoform, PI3Kδ-S, that is resistant to small molecule inhibitor idelalisib as defined by *in vitro* assays and mouse xenograft models. In contrast, the corresponding EA isoform (PI3Kδ-L) encodes a less aggressive isoform that is sensitive to idelalisib inhibition. Our present work now demonstrates that the short variant of *FGFR3* (*FGFR3-S*) encodes a more aggressive oncogenic signaling protein isoform, FGFR3-S, that is resistant to tyrosine kinase inhibitor dovitinib as defined by *in vitro* assays. In contrast, the corresponding EA isoform (FGFR3-L) encodes a less aggressive isoform that is sensitive to dovitinib inhibition.

Our discovery portends a genetic screening test for cancer patients that may be resistant to idelalisib or dovitinib therapy, which is particularly germane in hematologic and solid tumor malignancies, respectively. Idelalisib is FDA approved for relapsed chronic lymphocytic leukemia and relapsed non-Hodgkin lymphoma. Treatment resistance is observed in ~30-40% of patients, and the mechanism of resistance is currently unknown. Our hypothesis is that the short mRNA variant of *PIK3CD* may be playing a major role in treatment resistance. Clinical and molecular studies are currently underway to test this hypothesis. Dovitinib is currently being tested as an inhibitor of the FGFR3 in solid malignancies. The mRNA splicing pattern of the *FGFR3* gene is underappreciated at this juncture. Our hypothesis is that the short mRNA variant of *FGFR3* may be playing a major role in treatment resistance.

Impact on technology transfer:

Our findings that multiple variant proteins expressed in prostate cancer are resistant to small molecule inhibitors have sparked interests in companies that are investigating small molecule inhibitors of kinases involved in cancer progression. These companies are gaining an appreciation that alternative splicing in kinases can affect the sensitivity these signaling proteins to cancer therapeutic agents. Our findings highlights the potential importance of prescreening patients for their variant protein in order to prognosticate whether a particular small molecule inhibitor will be therapeutically efficacious.

Impact on society beyond science and technology:

Nothing to report

5. CHANGES/PROBLEMS

Changes in approach:

None

Actual or anticipated problems or delays:

None

Changes that had significant impact on expenditures:

None

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents:

None

6. PRODUCTS

Publications, conference papers, and presentations:

i. Wang B-D, Ceniccola K, Hwang S, Andrawis R, Horvath A, Freeman JA, Knapp S, Ching T, Garmire L, Patel V, Garcia-Blanco MA, Patierno SR, and **Lee NH** (2017) Alternative splicing promotes tumor aggressiveness and drug resistance in African American prostate cancer. ***Nature Communications*** 8:15921. PMCID:PMC5497057

ii. Wang Y, Freedman JA, Liu H, Moorman PG, Hyslop T, George DJ, **Lee NH**, Patierno SR, and Wei Q (2017) Associations between RNA splicing regulatory variants of stemness-related genes and racial disparities in susceptibility to prostate cancer. ***Int. J. Cancer*** 141:731-743. PMCID:PMC5512873

iii. Presentation at the Tenth AACR Conference The Science of Cancer Health Disparities, (pg 97), Atlanta, GA. Poster: Olender J., Wang B.-D., Nguyen, K., Lee N.H. (2017) "Novel FGFR3 splice variant increases oncogenic phenotype in African American prostate cancer"

iv. Presentation "Alternative splicing of FGFR3 as a mechanism for prostate cancer health disparities" at the 9th AACR Conference on The Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved, September 25 - 28, 2016

v. Manuscript in review. Freedman JA, Wang Y, Xuechan L, Liu H, Moorman P, George DJ, **Lee NH**, Hyslop T, Wei Q and Patierno SR (2018) Single nucleotide polymorphisms of stemness pathway genes predicted to regulate RNA splicing, microRNA and transcription are associated with prostate cancer survival. ***Int. J. Cancer*** (in review)

Website(s) or other Internet site(s):

None

Technologies or techniques:

None

Inventions, patent applications, and/or licenses:

i. **Lee NH**, Wang B-D (2015) United States Patent 14/639,219 - GWUV-001/01US 072905-2011. Companion diagnostics for cancer and screening methods to identify companion diagnostics for cancer based on splicing variants.

Other Products:

None

7. PARTICIPANTS & OTHER COLLABORATORS

Individuals working on this project:

Name:	Norman H Lee, PhD
Project Role:	PI
Nearest person month worked:	3
Contribution to project:	Direct and oversee entire project. Involved in experimental design and statistical analysis.

Name:	Bi-Dar Wang, PhD
Project Role:	Co-Investigator
Nearest person month worked:	3
Contribution to project:	Contributed to the cloning of variant cDNAs, in vitro, xenograft assays and cell-free protein purification and analysis.
Funding support:	Partially supported by GWU bridge support

Name:	Jacqueline Olender
Project Role:	PhD graduate student
Nearest person month worked:	12
Contribution to project:	Contributed to the cloning of variant cDNAs, in vitro and in vivo assays

Name:	Patricia Latham, MD
Project Role:	Co-Investigator
Nearest person month worked:	1

Contribution to project:	Animal necropsy and immunohistochemistry of xenografts

Change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period:

None

Other organizations were involved as partners:

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS

Nothing to report

9. APPENDICES

i. Wang B-D, Ceniccola K, Hwang S, Andrawis R, Horvath A, Freeman JA, Knapp S, Ching T, Garmire L, Patel V, Garcia-Blanco MA, Patierno SR, and **Lee NH** (2017) Alternative splicing promotes tumor aggressiveness and drug resistance in African American prostate cancer. ***Nature Communications*** 8:15921. PMCID:PMC5497057. Part of the published work was supported by W81XWH-13-1-0449.

nature COMMUNICATIONS

ARTICLE

Received 28 Apr 2016 | Accepted 15 May 2017 | Published xx xxx 2017

DOI: 10.1038/ncomms15921 OPEN

Alternative splicing promotes tumor aggressiveness and drug resistance in African American prostate cancer

Bi-Dar Wang[1,2], Kristin Ceniccola[1], SuJin Hwang[3], Ramez Andrawis[4], Anelia Horvath[1], Jennifer A. Freedman[5], Jacqueline Olender[1], Stefan Knapp[6,7], Travers Ching[8], Lana Garmire[8], Vyomesh Patel[9], Mariano A. Garcia-Blanco[10], Steven R. Patierno[5] & Norman H. Lee[1]

Clinical challenges exist in reducing prostate cancer (PCa) disparities. The RNA splicing landscape of PCa across racial populations has not been fully explored as a potential molecular mechanism contributing to race-related tumour aggressiveness. Here, we identify novel genome-wide, race-specific RNA splicing events as critical drivers of PCa aggressiveness and therapeutic resistance in African American (AA) men. AA-enriched splice variants of *PIK3CD*, *FGFR3*, *TSC2* and *RASGRP2* contribute to greater oncogenic potential compared with corresponding European American (EA)-expressing variants. Ectopic overexpression of the newly cloned AA-enriched variant, *PIK3CD-S*, in EA PCa cell lines enhances AKT/mTOR signalling and increases proliferative and invasive capacity *in vitro* and confers resistance to selective PI3Kδ inhibitor, CAL-101 (idelalisib), in mouse xenograft models. High *PIK3CD-S* expression in PCa specimens associates with poor survival. These results highlight the potential of RNA splice variants to serve as novel biomarkers and molecular targets for developmental therapeutics in aggressive PCa.

[1] Department of Pharmacology and Physiology, School of Medicine and Health Sciences, The George Washington University, Washington, District Of Columbia 20037, USA. [2] Department of Pharmaceutical Sciences, School of Pharmacy and Health Professions, University of Maryland Eastern Shore, Princess Anne, Maryland 21853, USA. [3] Department of Microbiology, Immunology and Tropical Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, District Of Columbia 20037, USA. [4] Department of Urology, School of Medicine and Health Sciences, The George Washington University, Washington, District Of Columbia 20037, USA. [5] Duke Cancer Institute and Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA. [6] Department of Clinical Pharmacology, University of Oxford, Oxford OX3 7BN, UK. [7] The Nuffield Department of Clinical Medicine, Structural Genomics Consortium, University of Oxford, Oxford OX3 7BN, UK. [8] Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. [9] Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892, USA. [10] Department of Biochemistry & Molecular Biology, The University of Texas Medical Branch at Galveston, Galveston, Texas 77555, USA. Correspondence and requests for materials should be addressed to N.H.L. (email: nhlee@gwu.edu).

Q1

Prostate cancer (PCa) is the most commonly diagnosed cancer and the second leading cause of cancer death among American men[1]. Striking population disparities in PCa risk and clinical outcome have been observed across racial groups. Notably, African American (AA) men exhibit 1.6-fold higher incidence and 2.4-fold higher mortality rates of PCa compared with European American (EA) men[2,3]. Socioeconomic factors remain a major component accounting for the PCa disparities between AA and EA populations[3,4]. However, higher mortality and recurrence rates are still observed in AA PCa even after adjustment of socioeconomic factors[5], suggesting that intrinsic biological differences also play a contributing role in PCa disparities[6–10].

Alternative splicing (AS) is a post-transcriptional process allowing for the generation of alternative mRNA transcripts that encode structurally and functionally disparate protein isoforms. Next-generation sequencing suggests that >90% of human genes undergo AS[11], and the resulting complexity in the transcriptome explains how ~20,000 protein-coding genes in the genome can lead to >250,000 distinct proteins in the proteome. Accumulating evidence indicates that alternative and/or aberrant splicing of precursor (pre)-mRNA plays an important but largely underappreciated role in cancers[12–15], including PCa[16]. For example, the *B-cell lymphoma 2-like 1* (*BCL2L1*) pre-mRNA is alternatively spliced into two variants, *Bcl-xS* and *Bcl-xL*, encoding protein isoforms with opposite biological effects[17]. Bcl-xS is a pro-apoptotic protein, while BcL-xL has anti-apoptotic properties conferring chemotherapy resistance in PCa cell line PC-3 (ref. 15) as well as castration-resistant xenograft growth[18]. Manipulation of splicing to decrease Bcl-xL and increase Bcl-xS levels has been shown to reduce tumour load[19]. *Fibroblast growth factor receptor 2* (*FGFR2*) pre-mRNA also undergoes AS, where *FGFR2-IIIb* is predominately expressed in epithelial cells and *FGFR2-IIIc* is primarily associated with

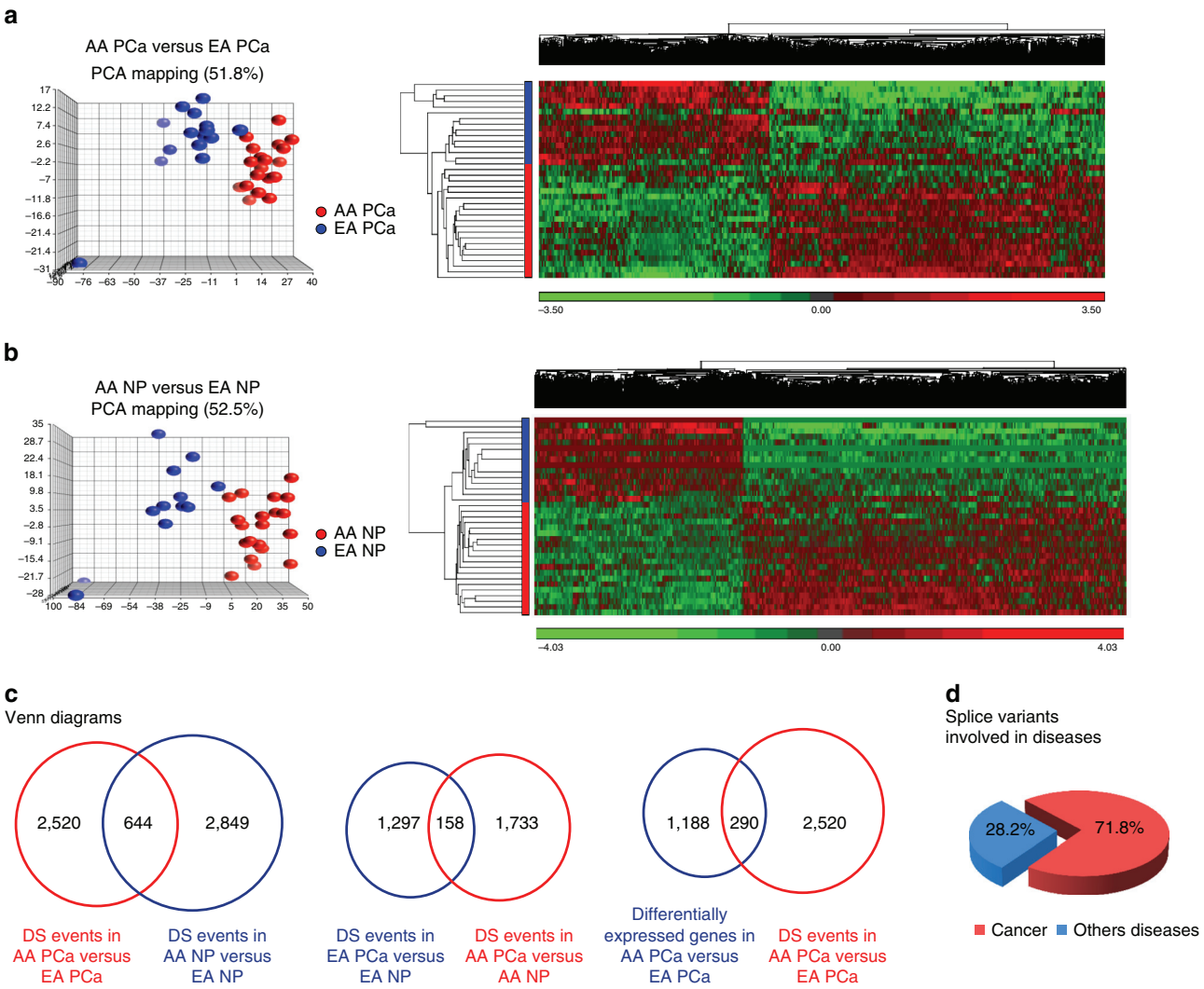

Figure 1 | Differential alternative splicing events in AA PCa compared with EA PCa and AA NP compared with EA NP specimens. (**a**) Principal component analysis (PCA) plot and two-dimensional (2D) clustergram depicting 3,112 significant differentially expressed exons in 20 independent AA PCa versus 15 independent EA PCa specimens. (**b**) PCA plot and 2D clustergram depicting 3,384 significant differentially expressed exons in 20 AA NP versus 15 EA NP specimens. AA and EA specimens are represented by red and blue circles/bars, respectively. Rows represent specimens and columns represent exons in hierarchical clustergrams. Log2 expression values of exons were subjected to 2D unsupervised hierarchical clustering using average linkage method and Euclidean distance. (**c**) Venn diagrams of DS events in AA PCa versus EA PCa and AA NP versus EA NP, DS events in EA PCa versus EA NP and AA PCa versus AA NP and differentially expressed genes in AA PCa versus EA PCa and DS events in AA PCa versus EA PCa. (**d**) A majority of the genes with DS events in AA PCa versus EA PCa were functionally associated with cancer. The top three 'other diseases' were gastrointestinal disease, organismal injury and abnormalities and reproductive system disease.

epithelial-to-mesenchymal transition of PCa cells[20]. Another example is the *TMPRSS2-ERG* gene fusion commonly found in PCa and associated with poor clinical outcome[21,22]. In a comparative study of *TMPRSS2-ERG* variants ectopically overexpressed in prostatic epithelial cells, variants containing a 72 base exon (*+72*) mediate increased cell proliferation and invasion[14]. Androgen receptor (AR) signalling is critically associated with PCa growth[23] and splice variant *AR-V7* is overexpressed in hormone-refractory PCa, being correlated with poor patient survival and higher recurrence rates[24,25].

Despite the significance of alternative/aberrant splicing in PCa progression irrespective of race, the occurrence of race-specific/-enriched PCa splicing events and a causal relationship between these events and observed PCa disparities remains unexplored. For example, it is unclear whether *FGFR2-IIIc*, *TMPRSS2-ERG+72*, *AR-V7* and/or other as yet undiscovered variants associated with more aggressive PCa might be predominantly or selectively expressed in AA PCa, thus contributing to PCa disparities. In addition, it is unknown whether differences in mRNA splicing along racial/population lines occur in only a limited number of genes or more globally across the transcriptome. If the latter, it will be important to ascertain whether these genome-wide, differential splicing (DS) events are overrepresented within specific gene ontologies (that is, proto-oncogenes, tumour suppressor genes). Lastly, assessment of the functional consequences of any race-specific (or enriched) splicing events will provide critical further insight into the genetic/molecular mechanisms underlying PCa disparities. To this end, we have applied a functional genomics approach to address these questions. Our results underscore the leveraging of population differences in tumour biology to discover novel splice variants that will likely serve as novel biomarkers and/or molecular targets for developmental therapeutics against aggressive AA PCa, identify previously hidden splice variants encoding oncogenic signalling proteins resistant to small-molecule inhibitors (SMIs), and assimilate splice variant information for prognostication of cancer aggressiveness and/or therapeutic responsiveness.

Results

Genome-wide DS events in AA versus EA PCa. A total of 35 PCa (20 AA/15 EA) and 35 patient-matched normal prostate (NP) specimens (20 AA/15 EA) derived from chemo-/hormone-/radiation-naive patients were interrogated using the Affymetrix Human Exon 1.0 ST GeneChip to assess DS events. Gleason scores of PCa specimens (range 6–8) and patient ages (range 49–81 years) were not significantly different between AA and EA cohorts ($P>0.05$, Fisher's exact test). In AA PCa versus EA PCa and AA NP versus EA NP, the significant differentially expressed exons (Fig. 1a,b) could be modelled using the AS analysis of variance (ANOVA) approach[26] into 2,520 and 2,849 DS events, respectively (Supplementary Data 1 and Supplementary Fig. 1). As depicted in the Venn diagram (Fig. 1c), 1,876 genes (2,520 minus 644) exhibited DS events unique to AA PCa versus EA PCa, 2,205 differentially spliced genes (2,849-644) were unique to AA NP versus EA NP and 644 DS events were in common (that is, DS events preexisting in AA NP versus EA NP and preserved in AA PCa versus EA PCa). Examples of genes with preexisting DS events included *PIK3CD*, *ITGA4* and *MET*, while *RASGRP2*, *NF1* and *BAK1* are examples of differentially spliced genes occurring only in AA PCa versus EA PCa. In EA PCa versus EA NP and AA PCa versus AA NP, the significant differentially expressed exons (Supplementary Fig. 2) could be modelled into 1,297 and 1,733 DS events, respectively (Fig. 1c). Presumably, a subset of 1,575 genes (1,733 − 158) with DS events unique to AA PCa may contribute to PCa disparities. Examples in this category included *FGFR3* and *TSC2* (Supplementary Data 1). On the other hand, a subset of 158 genes with DS events in common to both AA and EA PCa may contribute to PCa progression regardless of race (Fig. 1c). Consistently, such genes included *TMPRSS2* and *AR* (Supplementary Data 1c,d). Analysis of the exon array data employing both gene-wise[9,10] and AS ANOVA modelling approaches[26] identified 898 genes (1,188 − 290) that were differentially expressed but not exhibiting DS in AA PCa versus EA PCa, and 2,230 (2,520 − 290) genes undergoing DS but not differential expression (for example, level of variant 'A' for gene 'X' in AA PCa equivalent to variant 'B' for gene 'X' in EA PCa; Fig. 1c).

Prevalence of DS events in cancer-associated pathways. We categorized genes undergoing DS in AA PCa versus EA PCa based on molecular function, gene ontology and disease association. Relevant cancer-related ontologies included cell growth and proliferation, cell death and survival, cellular movement, cell adhesion and DNA damage/repair (*P* values ranged from 6.54×10^{-12} to 1.88×10^{-2}, Fisher's exact test; Supplementary Data 2). Notably, a large fraction (1,816 out of 2,520, 71.8%) of the differentially spliced genes were discovered to be over-represented across multiple cancers, including colorectal, renal, breast, brain, lung, stomach, prostate and haematologic cancers (*P* values ranged from 1.43×10^{-9} to 1.96×10^{-2}, Fisher's exact test; Fig. 1d and Supplementary Data 2). There was an unexpected skewing in the distribution of in-frame versus out-of-frame exon skipping events in cancer-related genes, where in-frame events were significantly favoured in AA over EA PCa specimens ($P<0.05$, Fisher's exact test; Supplementary Table 1). This finding was in line with an overall significant preference for in-frame events across all genes (cancer-related and noncancer-related) in AA PCa specimens (Supplementary Table 1). In the case of noncancer-related genes only, there was no significant skewing of in-frame distribution events between AA versus EA PCa (Supplementary Table 1).

We also examined the distribution of DS events across cell signalling pathways. There was a striking significant over-representation of DS events in multiple oncogenic signalling pathways, including epidermal growth factor (EGF), vascular endothelial growth factor (VEGF), phosphatase and tensin homolog (PTEN), phosphatidylinositol-3-kinase (PI3K)/AKT, extracellular signal–regulated kinase/mitogen-activated protein kinase (ERK/MAPK) and nuclear factor-κB (NF-κB) signalling (Fisher's exact test, *P* values ranged from 0.00126 to 0.02089; Supplementary Fig. 3 and Supplementary Data 3). Interestingly, many of these same pathways are known to be mutated based on earlier cancer genome sequencing studies[27–29]. A composite oncogenic signalling pathway comprising DS events found in AA PCa versus EA PCa is depicted in Fig. 2. Taken together, our data provide strong evidence that DS events may play a critical role in PCa disparities.

Validation of AS variants in AA versus EA PCa. We proceeded to validate a subset of both proto-oncogenes and tumour suppressor genes with DS events in our composite cancer signalling pathway, including *PIK3CD*, *FGFR3*, *TSC2*, *ITGA4*, *MET*, *NF1*, *BAK1*, *ATM* and *RASGRP2* (Fig. 2). Real-time PCR (RT-PCR) was performed on RNA samples obtained from AA and EA PCa specimens originally interrogated by the exon arrays. Primer pairs or trios were designed for RT-PCR to amplify simultaneously multiple variants of each gene (Fig. 3a). As shown in Fig. 3b, AA PCa specimens contained both *PIK3CD* long (*PIK3CD-L*, including exon 20) and short (*PIK3CD-S*, missing exon 20) variants, whereas EA PCa samples predominately

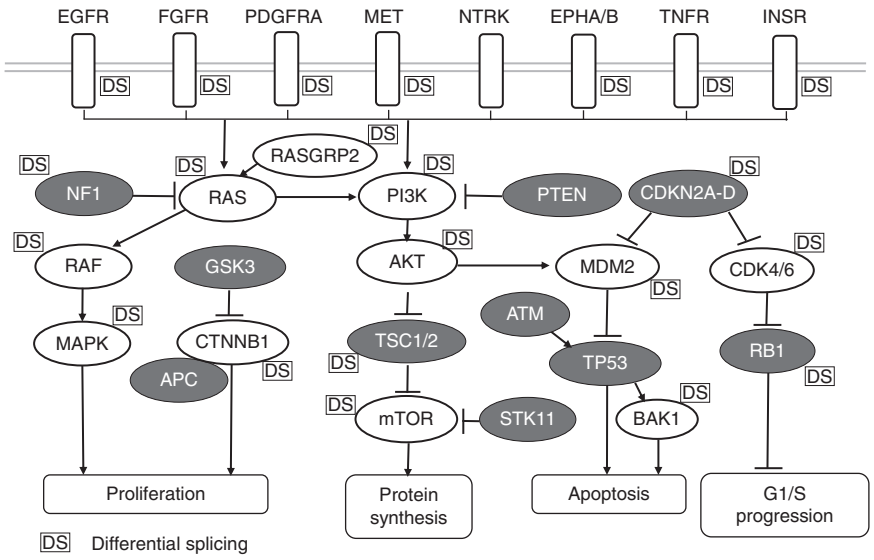

Figure 2 | Representation of differential splicing events in a composite oncogenic signalling pathway. DS events in AA PCa versus EA PCa were frequently detected in multiple oncogenic signalling pathways. Depicted pathway is a composite of PI3K/AKT/mTOR, RAS/RAF/MAPK, CDK/RB1, MDM2/TP53 and WNT/GSK3/CTNNB1/APC signalling. Open figures indicate oncogenes and closed figures indicate tumour suppressor genes. 'DS' indicates that a differential alternative splicing event was detected for a particular oncogene or tumour suppressor gene.

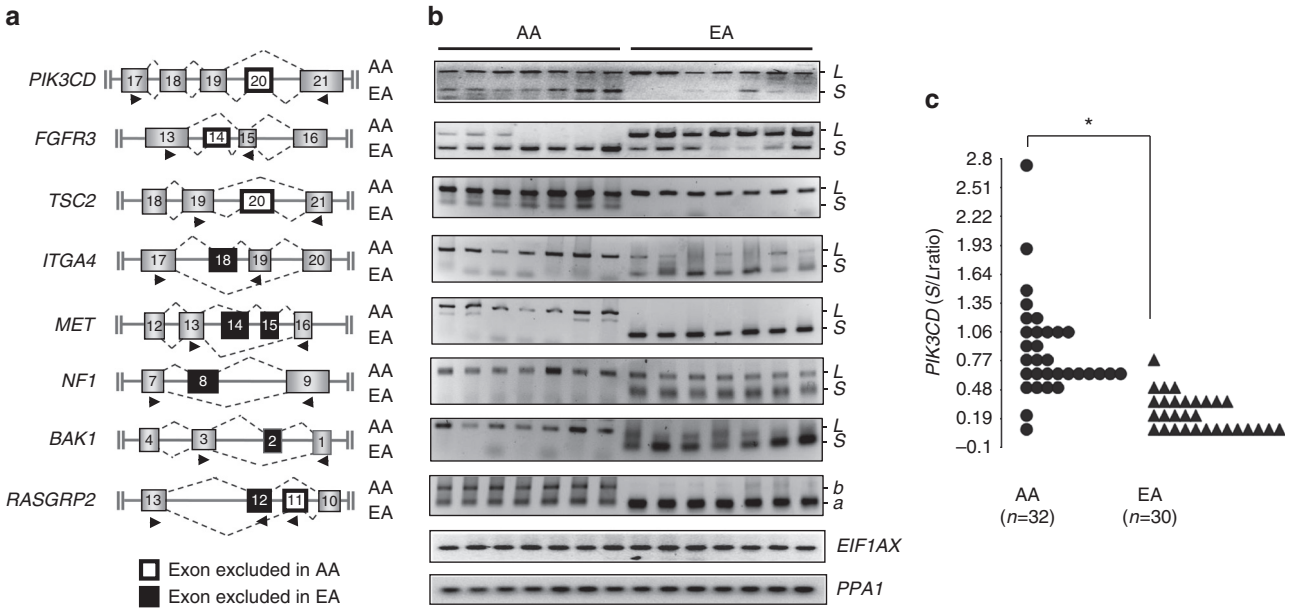

Figure 3 | Validation of differential alternative splicing of oncogenes and tumour suppressor genes in AA versus EA PCa specimens. (**a**) Schematic representation of DS events within the indicated genes in AA PCa (top dashed lines connecting exons) and EA PCa (bottom dashed lines) based on alternative splicing ANOVA model. Closed arrowheads below exons represent primer location in qRT-PCR validation of alternatively spliced transcripts. (**b**) Representative RT-PCR results validating race-specific/-enriched variant transcripts in either AA PCa or EA PCa specimens. Shown are the RT-PCR results for the AA-specific/-enriched variants *PI3KCD-S*, *FGFR3-S*, *TSC2-S*, *ITGA4-L*, *MET-L*, *NF1-L*, *BAK1-L* and *RASGRP2-b*; and EA-specific/-enriched variants *PIK3CD-L*, *FGFR3-L*, *TSC2-L*, *ITGA4-S*, *MET-S*, *NF1-S*, *BAK1-S* and *RASGRP2-a*. Each lane represents an RT-PCR result from an independent PCa specimen that was also interrogated in exon array experiments. RT-PCR of *EIF1AX* and *PPA1* transcripts served as loading controls. The qRT-PCR results are summarized in Supplementary Fig. 4. Unprocessed RT-PCR images are shown in Supplementary Fig. 11. (**c**) Additional validation and quantification of the ratio of the AA-enriched *PI3KCD-S* (short) variant and the race-independent *PI3KCD-L* (long) variant in a separate cohort of PCa patient specimens. RNA was isolated from $n=32$ AA PCa and $n=30$ EA PCa specimens and subjected to qRT-PCR. Shown is a plot of the ratio of *S/L*. *EIF1AX* and *PPA1* transcripts served as internal normalization controls. $^*P<0.05$ using two-sided Student's *t*-test. Variance was similar among groups being compared.

expressed *PIK3CD-L*. Hence, the RT-PCR results were in agreement with the exon array data, indicating the presence of an AA-enriched *PIK3CD-S* variant. Analogous findings were obtained where either a short or long variant of each gene was confirmed by RT-PCR to be enriched or uniquely expressed in AA (*TSC2-S*, *ITGA4-L*, *MET-L*, *BAK1-L*) or EA PCa (*FGFR3-L*, *ITGA4-S*, *MET-S*, *NF1-S*, *BAK1-S*) (Fig. 3b; Supplementary Fig. 4 for quantitative RT-PCR (qRT-PCR) results from $n=22$–25 AA and $n=21$–24 EA PCa specimens). Exon array data also revealed two alternative *RASGRP2* transcripts with apparent mutually

exclusive exon skipping events (Fig. 3a and Supplementary Fig. 1d). RT-PCR validation likewise confirmed that a *RASGRP2-b* variant (exon 11 excluded) was exclusively expressed in AA PCa, while a *RASGRP2-a* variant (exon 12 excluded) was enriched in EA PCa (Fig. 3b and Supplementary Fig. 4). We were unable to validate DS of *ATM*, whereas two additional genes (*GSK3A* and *EPHA1*) identified by exon arrays as not undergoing DS were confirmed by RT-PCR. In summary, there was strong agreement (10/11, 91%) between exon array and RT-PCR results, thus providing an internal quality metric to our global DS analysis of AA and EA PCa (see Methods for additional metrics).

The race-dependent expression of *PIK3CD* variants was particularly interesting owing to recent findings implicating PI3Kδ (p110δ) kinase activity in haematologic malignancies as well as other cancer types[30–33]. In a separate cohort of PCa specimens obtained from 32 AAs (age range 52–76 years, Gleason score range 6–8) and 30 EAs (age range 50–82 years, Gleason score range 6–8; not significantly different from AA, Fisher's exact test, $P>0.05$), quantitative RT-PCR validation was performed reaffirming significantly higher levels of *PIK3CD-S* relative to *PIK3CD-L* in AA versus EA PCa specimens (Fig. 3c). Given the robustness and potential significance of these findings, subsequent *in vitro* and *in vivo* studies centred on the *PIK3CD* variants, as described below.

Molecular cloning of *PIK3CD* splice variants. The AA-enriched *PIK3CD-S* variant has never before been described in the literature nor the UCSC (University of California, Santa Cruz) (genome.ucsc.edu) or Ensembl Genome Browser (www.ensembl.org). Consequently, we cloned the full-length versions of *PIK3CD-S* from AA PCa cell line MDA PCa 2b, and *PIK3CD-L* from MDA PCa 2b as well as EA PCa cell lines VCaP and LNCaP using standard molecular approaches (5′- and 3′-RACE (rapid amplification of cDNA ends)[34]). We likewise cloned matching *PIK3CD-S* and *PIK3CD-L* variants from PCa patient specimens. Supplementary Fig. 5 schematically depicts the full-length clones of *PIK3CD-L* (comprising a total of 24 exons) along with three different AA *PIK3CD-S* variants (variant excluding exon 8, exon 20 or both exons 8 and 20) and one AA large deletion variant of the *PIK3CD* gene. Interestingly, exclusion of exon 8 eliminates a 30-amino acid segment situated between the Ras-binding and C2 domains, while exclusion of exon 20 deletes a 56-amino acid segment located in the catalytic domain of PI3Kδ. In subsequent functional studies involving ectopic overexpression of the short variant (see below), we concentrated our efforts on the variant missing exon 20 given the possibility that kinase activity may be affected.

***PIK3CD-S* isoform augments invasion and proliferation**. We hypothesized that the splice variants specific or enriched in AA PCa may contribute to a more aggressive oncogenic phenotype. To test this, we designed exon-specific and exon junction-specific short interfering RNAs (siRNAs) to target *PIK3CD-L* and *PIK3CD-S*, respectively, in EA and AA PCa cell lines and examined the functional consequences of these knockdowns on cell proliferation and invasion. A similar strategy was applied to investigate the biological significance of the variants of *FGFR3*, *TSC2* and *RASGRP2*. VCaP and MDA PCa 2b cells were used as population-specific PCa models, as these two cell lines represent bone metastases derived from castration-resistant EA and AA PCa patients, respectively[35,36]. Transfection of VCaP cells with exon 20-specific siRNA (siP_{20}) successfully knocked down *PIK3CD-L* expression by >8-fold compared with nonsense siRNA (Fig. 4a, left panel), resulting in a significant loss of proliferative and invasive function in VCaP cells (Fig. 4b, left). Conversely, in MDA PCa 2b cells, a >5-fold knockdown of *PIK3CD-L* increased the ratio of *PIK3CD-S/PIK3CD-L* expression by nearly twofold (Fig. 4a, right panel; 1.88 *S/L* ratio for nonsense versus 3.46 *S/L* ratio for siP_{20}-transfected cells), and this 'enrichment' of AA-enriched *PIK3CD-S* subsequently enhanced proliferation and invasion of the AA cell line (Fig. 4a, right panel). Moreover, MDA PCa 2b cells exhibited significantly higher basal invasive and proliferative capacities compared with VCaP cells (proliferation and invasion of siNS-transfected MDA PCa 2b versus siNS-transfected VCaP; Fig. 4b, left and right panels). To further evaluate the functional impact of *PIK3CD-S* expression on cell proliferation and invasion, the EA and AA PCa cell lines were transfected with siP_j (siRNA specifically targeting the junction of exons 19 and 21). Transfection of siP_j had no effect on VCaP proliferation and invasion, as expected since this EA line does not significantly express *PIK3CD-S* (Fig. 4a,b; right panels). On transfection of MDA PCa 2b cells with siP_j, *PIK3CD-S* expression was significantly knocked down (Fig. 4a, right), resulting in a loss of cell proliferation and invasion (Fig. 4b, right). Taken together, these results suggest that *PIK3CD-S* is the more aggressive variant, promoting PCa proliferation and invasion to a greater extent than *PIK3CD-L*.

Several additional exon-specific siRNAs were designed to test whether other AA-specific/-enriched splice variants also functionally contribute to greater PCa aggressiveness. SiRNAs targeting exon 14 (*siFGFR3-ex14*), exon 20 (*siTSC2-ex20*) and exon 11 (*siRASGRP2-ex11*) were used to selectively suppress expression of *FGFR3-L*, *TSC2-L* and *RASGRP2-a* variants (predominately expressed in EA), respectively. Upon siRNA-mediated knockdown of *FGFR3-L*, *TSC2-L* or *RASGRP2-a* (Supplementary Fig. 6a–c, top panels) in MDA PCa 2b cells, the expression ratios of *FGFR3-S/FGFR-L*, *TSC2-S/TSC2-L* and *RASGRP2-b/RASGRP2-a* increased and correlated with augmented invasive and/or proliferative capacity of the AA-derived MDA PCa 2b cells (Supplementary Fig. 6a–c, bottom panels). Collectively, our *in vitro* studies strongly suggest that the AA-enriched splice variants *PIK3CD-S*, *FGFR3-S*, *TSC2-S* and *RASGRP2-b* promote PCa aggressiveness.

***PIK3CD-S* isoform promotes activation of AKT/mTOR signalling**. As PI3K plays a central role in the PI3K/AKT/mammalian target of rapamycin (mTOR) signalling pathway, we examined the ability of different PI3Kδ isoforms (encoded by *PIK3CD-L* and *PIK3CD-S*) to activate downstream signalling components within this pathway. SiRNA (siP_{20})-mediated knockdown of *PIK3K-L* expression (confirmed by qRT-PCR) led to a drastic decrease in AKT phosphorylation at Thr308 and Ser473 while moderately decreasing (approximately twofold) phosphorylation of mTOR in EA VCaP cells (Fig. 4c, top panel). In contrast, knockdown of *PIK3CD-L* in AA MDA PCa 2b cells (resulting in an increased *PIK3CD-S/-L* ratio confirmed by qRT-PCR) led to a sizable increase (two- to threefold) in phosphorylation of AKT, mTOR and ribosomal protein S6 (S6) (Fig. 4c, top panel).

In parallel experiments, siRNA (siP_j)-mediated knockdown of *PIK3K-S* (confirmed by qRT-PCR) in MDA PCa 2b cells resulted in a drastic decrease in the phosphorylation status of AKT, mTOR and S6 (Fig. 4c, bottom panel). As expected, treatment of VCaP cells with the siRNA siP_j had negligible effects on AKT, mTOR and S6 phosphorylation, as this EA line does not significantly express *PIK3CD-S*. Taken together, the distinct phosphorylation patterns of AKT, mTOR and S6 in AA and EA PCa cell lines upon selective knockdown of either *PIK3CD-L* or *PIK3CD-S* again suggested that *PIK3CD-S* is the more aggressive variant, promoting oncogenic signalling.

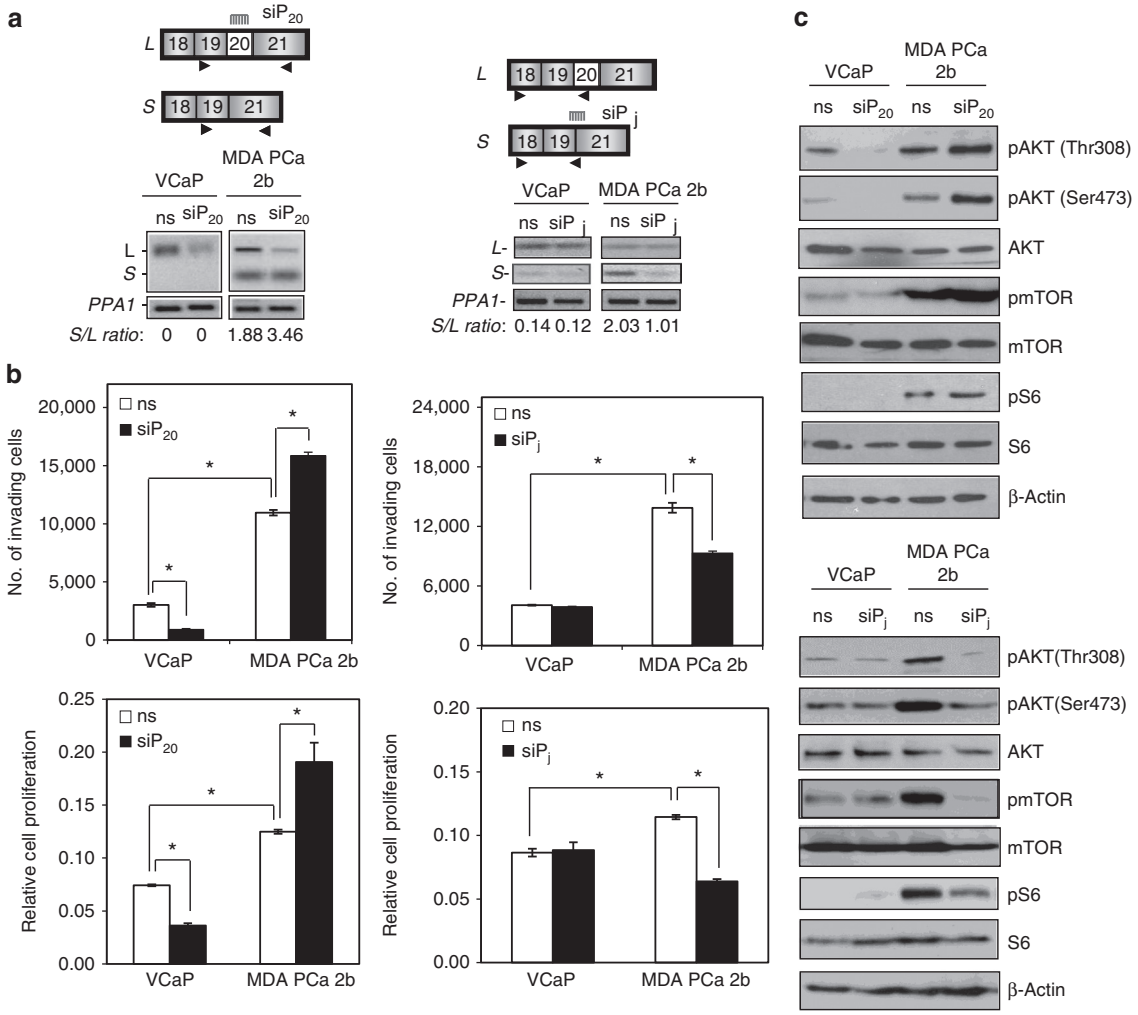

Figure 4 | Selective knockdown of the AA-enriched *PIK3CD-S* variant or race-independent *PIK3CD-L* variant has opposing effects on proliferation, invasion and AKT/mTOR signalling in AA PCa cells. (**a**) RT-PCR of *PIK3CD-L* and *PIK3CD-S* variants following knockdown in the EA PCa cell line VCaP and AA PCa cell line MDA PCa 2b. Specific knockdown of the *PIK3CD-L* variant (left panel) was accomplished with an exon 20-targeting siRNA (siP_{20}), while knockdown of the *PIK3CD-S* variant (right panel) was achieved with an siRNA targeting the region spanning exons 19 and 21 (siP_j). Closed arrowheads below exons represent primer location for qRT-PCR validation of alternatively spliced transcript knockdowns. Knockdown efficiency of siP_{20} and siP_j siRNAs was determined by the *S/L* ratio derived from RT-PCR reactions; ns, nonsense siRNA treatment. Representative images of $n=4$ independent knockdown experiments. (**b**) Proliferation and invasion of VCaP and MDA PCa 2b following knockdown of the *PIK3CD-S* (left panel) or *PIK3CD-L* variant (right panel). Data presented as mean ± s.e.m. from at least three independent experiments for each treatment group. *$P<0.05$ by ANOVA with *post hoc* Tukey. Variance was similar among groups being compared. (**c**) Western blot analysis of AKT/mTOR signalling following knockdown of the *PIK3CD-L* (top panel) or *PIK3CD-S* variant (bottom panel) in VCaP and MDA PCa 2b. Level of AKT, mTOR and ribosomal S6 kinase activities is reflected by the amount of phospho-AKT (pAKT), phospho-mTOR (pmTOR) and phospho-ribosomal protein S6 (pS6) immunoblotting, respectively. β-Actin served as loading control. Representative images from at least three independent western blot experiments.

PI3Kδ-S isoform is resistant to SMIs. We tested whether pharmacological inhibition of PI3Kδ isoforms represented a potential strategy for ameliorating PCa aggressiveness. CAL-101, an SMI specific for PI3Kδ (refs 37–39), was employed to assess its inhibitory effects on oncogenic signalling and proliferation in EA VCaP and PC-3 cells that stably overexpressed the His-tagged PI3Kδ-S (excluding exon 20) or PI3Kδ-L isoform (including exon 20). Equivalent levels of PI3Kδ isoform expression in each cell line was confirmed by western blot with a His tag antibody (Fig. 5a). In both EA cell lines, ectopic overexpression of PI3Kδ-S was associated with a two- to threefold greater phosphorylation of AKT and S6 compared with ectopic overexpression of PI3Kδ-L (Fig. 5a, absence of CAL-101 treatment). CAL-101 ($50\,mg\,kg^{-1}$) induced a significant reduction in basal AKT and S6 phosphorylation (Fig. 5a) and a dose-dependent inhibition of proliferation in both EA cell lines overexpressing the PI3Kδ-L variant (Fig. 5b). In contrast, CAL-101 ($50\,mg\,kg^{-1}$) had negligible effects on inhibiting basal AKT signalling in EA PCa cell lines overexpressing PI3Kδ-S, as phosphorylation states of AKT and S6 were comparable to vehicle-treated cells (Fig. 5a). In agreement, 5-bromodeoxyuridine labelling assays demonstrated that proliferation of VCaP and PC-3 cells ectopically overexpressing PI3Kδ-S was greater than cells overexpressing PI3Kδ-L (Supplementary Fig. 7). Moreover, PI3Kδ-S-overexpressing VCaP and PC-3 cells were not effectively inhibited by CAL-101 treatment; only at extreme doses of CAL-101 ($\geq 30\,\mu M$) was proliferative activity in PC-3 cells significantly impaired (Fig. 5b). In contrast, the AKT inhibitor MK-2206 (ref. 40) dose-dependently decreased proliferation in both PI3Kδ-S- and PI3Kδ-L-overexpressing VCaP and PC-3 cells (Fig. 5c). These results suggest that PI3Kδ-S-stimulated proliferation is resistant to CAL-101 inhibition in sharp contrast to PI3Kδ-L;

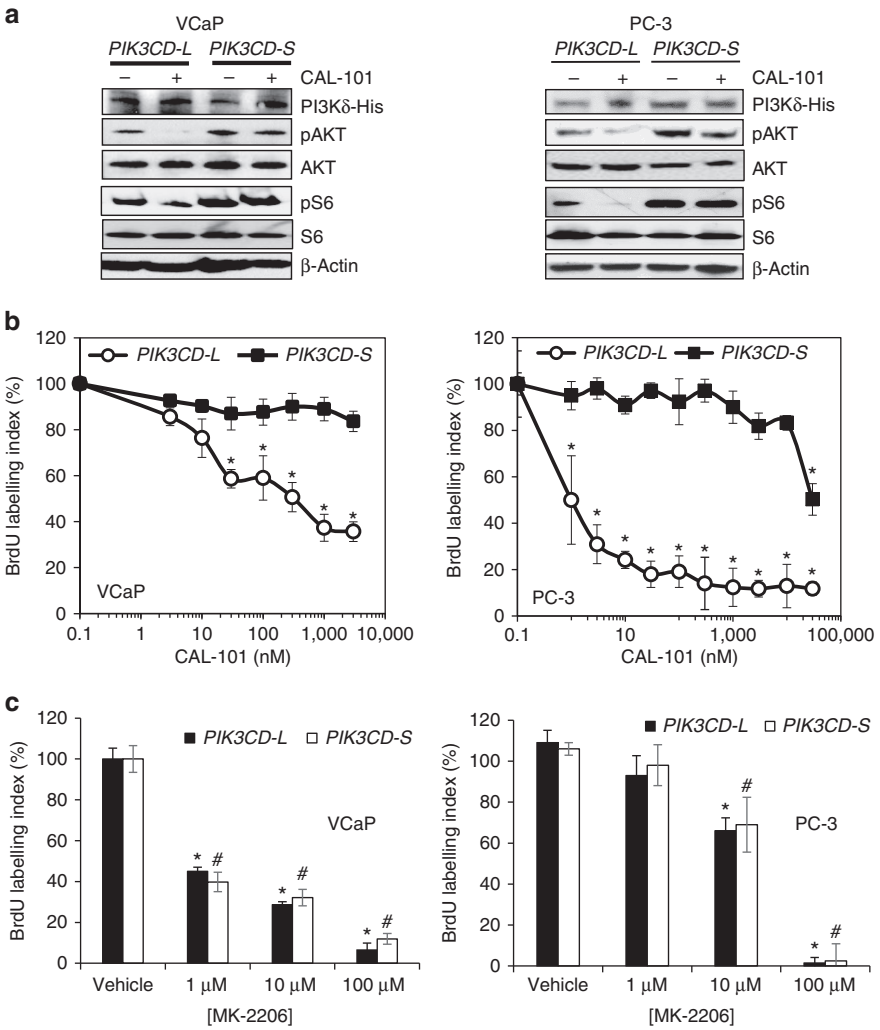

Figure 5 | PI3Kδ-S but not PI3Kδ-L is resistant to small-molecule inhibition of PI3K/AKT/mTOR signalling and proliferation. (**a**) Assessment of PI3K/AKT/mTOR signalling following treatment with vehicle (saline) or CAL-101 (100 nM, 24 h). PI3K/AKT/mTOR signalling was assessed by western blot analysis with phospho-antibodies to AKT (pAKT) and S6 ribosomal protein (pS6). β-Actin served as a loading control. His-tag antibody was used to demonstrate equal expression of His-tagged variant PI3Kδ protein in stably transfected cell lines. Representative images from at least three independent western blot experiments. Unprocessed western images shown in Supplementary Fig. 12. (**b**) Proliferation in VCaP and PC-3 cells stably overexpressing the *PIK3CD-S* variant or *PIK3CD-L* variant following treatment with vehicle (saline) or selective PI3Kδ small molecule inhibitor CAL-101 (24 h). Data presented as mean ± s.e.m. from at least four independent experiments for each treatment group. *Significantly different from -S variant, $P<0.05$ by ANOVA with Dunnett's *post hoc* test. (**c**) Treatment of *PIK3CD* variant-overexpressing cells with vehicle (saline) or selective AKT small molecule inhibitor MK-2206 (24 h). Proliferation was assessed using a 5-bromodeoxyuridine (BrdU) labelling assay. Data presented as mean ± s.e.m. from at least four independent experiments for each treatment group. * Or #Significantly different from corresponding vehicle control, $P<0.05$ by ANOVA with Dunnett's *post hoc* test. Variance was similar among groups being compared.

while inhibition of AKT, which is downstream of PI3Kδ-S, effectively blocked proliferation.

To examine the effects of *PIK3CD* splice variants on tumour growth *in vivo*, we subcutaneously injected 2×10^6 PC-3 cells stably overexpressing equivalent amounts of the PI3Kδ-L or PI3Kδ-S (missing exon 20) isoform into the left hind flank of nonobese diabetic–severe combined immunodeficient (NOD-SCID) mice. Mice harbouring PI3Kδ-L-expressing or PI3Kδ-S-expressing PC-3 cell xenografts were administrated either vehicle (phosphate-buffered saline) or CAL-101 ($50\,mg\,kg^{-1}$) by daily intraperitoneal (i.p.) injection. CAL-101 treatment for 30 days significantly reduced the growth of PI3Kδ-L-expressing xenografts compared with the vehicle treatment (Fig. 6a,b). In contrast, mice with xenografts of PI3Kδ-S-expressing cells had negligible suppression of their xenograft growth following CAL-101 treatment compared with vehicle-treated animals (Fig. 6a,b).

We further examined the inhibitory effects of CAL-101 on PI3Kδ isoforms in an *in vivo* tumour metastasis model. The 1×10^6 *PIK3CD-L*- or *PIK3CD-S*-overexpressing PC-3 cells were injected into the tail vein of NOD-SCID mice, and animals were subsequently administrated with vehicle or CAL-101 ($50\,mg\,kg^{-1}$) via i.p. injection (3 times per week). After 8 weeks, vehicle-treated mice carrying PI3Kδ-L-overexpressing cells developed prominent tumour metastases in the lungs (Fig. 6c,d), while the CAL-101 treatment group exhibited a >50% reduction ($P<0.05$) of metastases (Fig. 6c,d). In comparison, CAL-101 treatment failed to significantly inhibit tumour metastases in mice harbouring PI3Kδ-S-expressing cells (Fig. 6c,d). Noteworthy, the size of lung metastases (average area of nodules) in mice

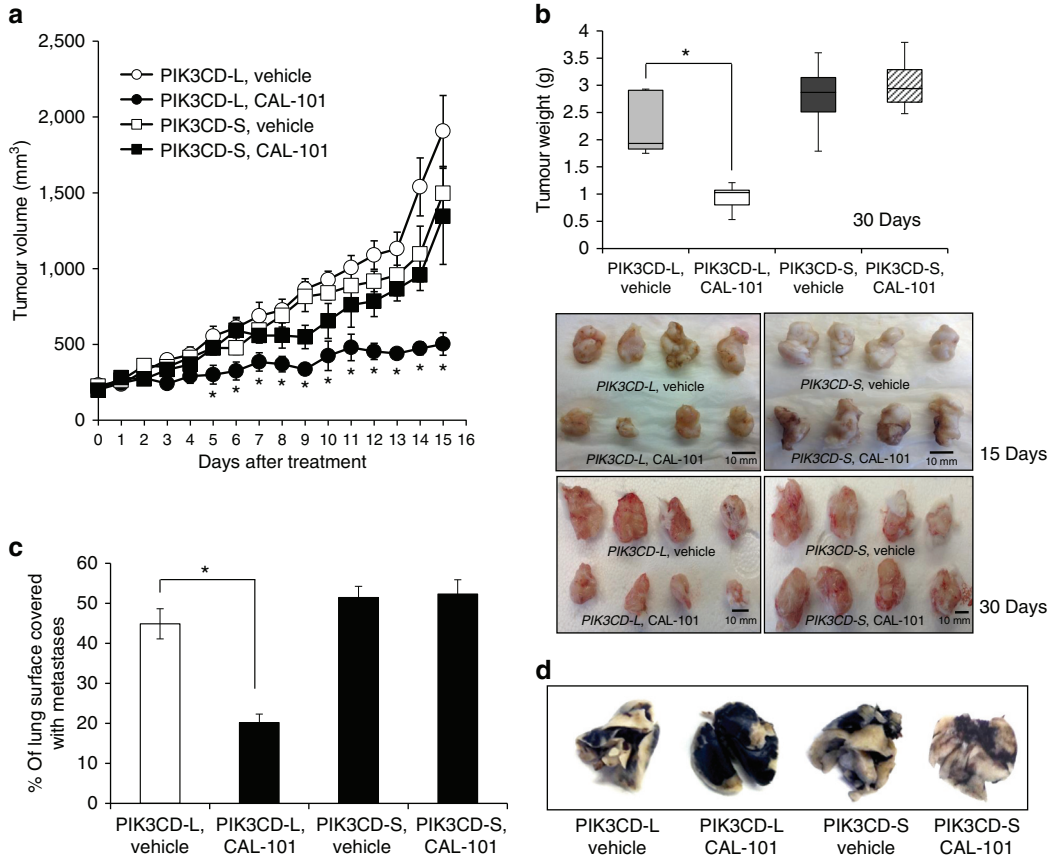

Figure 6 | PI3Kδ-S but not PI3Kδ-L is resistant to small-molecule inhibition of xenograft growth and metastasis. (**a**) Growth of PC-3 cells stably overexpressing the *PIK3CD-S* variant in NOD-SCID mice is resistant to CAL-101 treatment (50 mg kg^{-1} i.p. 3 times a week). In contrast, growth of PC-3 cells stably overexpressing the *PIK3CD-L* variant is sensitive to CAL-101 treatment. Data represent the mean tumour size ± s.e.m. of $n=10$ independent mice for each treatment group at each time point. *Significantly different from saline-treated group, $P<0.05$ by ANOVA with *post hoc* Tukey. Variance was similar among groups being compared. (**b**) Tumour weights and gross morphology of tumour xenografts from **a**. Box-and-whisker plot represents mean xenograft weight in mice after 30-day vehicle or CAL-101 treatment. *$P<0.05$ by ANOVA with Dunnett's *post hoc* test; $n=10$ independent mice for each treatment group. Variance was similar among groups being compared. (**c**) Quantification of lung metastases in NOD-SCID mice. PC-3 cells stably overexpressing *PIK3CD-L* or *PIK3CD-S* were injected into the tail vein of NOD-SCID mice treated with vehicle or CAL-101 (50 mg kg^{-1} i.p. 3 times a week). After 8 weeks, lungs were collected and stained with India ink and bleached with Fekete's solution for visualization of metastatic nodules (white-coloured areas). Data presented as mean ± s.e.m. of $n=10$ for each treatment group. *$P<0.05$ by ANOVA with Dunnett's *post hoc* test. Variance was similar among groups being compared. (**d**) Representative India ink-stained lungs from treatment groups analysed in **c**.

harbouring PI3Kδ-S-overexpressing cells was slightly greater (~15%) compared with animals with PI3Kδ-L-overexpressing cells, although not statistically significant ($P>0.05$). Taken together, the *in vitro* and *in vivo* functional studies suggest that SMIs such as CAL-101 (competitive ATP binding inhibitors[39,41]) may be ineffective against the PI3Kδ-S isoform in AA PCa.

Cell-free PI3Kδ isoform kinase assay. The consequence of excluding exon 20 (168 bp) in the *PIK3CD-S* variant is an in-frame deletion of 56 amino acids (residues 810–865) in the catalytic domain of the PI3Kδ-S isoform (Fig. 7a). To gain further insight into the functional differences between PI3Kδ isoforms, the interaction of PI3Kδ-L and -S with regulatory subunit p85α was investigated. Whole cell lysates from transfected PC-3 cells overexpressing p85α and either His-tagged PI3Kδ-S or PI3Kδ-L were subjected to western analysis, demonstrating that each cell line expressed equivalent levels of their respective PI3Kδ isoform as well as equal p85α expression (Fig. 7b, left panel). Interestingly, co-immunoprecipitation (co-IP) of the PI3Kδ/p85α complex from whole cell lysates using an anti-His antibody demonstrated that p85α bound with three- to fourfold greater proficiency to PI3Kδ-L compared with p85α binding to PI3Kδ-S (Fig. 7b, right panel, column E). Binding proficiency was inversely correlated with PI3Kδ isoform kinase activity (Fig. 7c, right panel).

Next, PI3Kδ isoforms were purified from the lysates of PC-3 cells overexpressing either His-tagged PI3Kδ-S or PI3Kδ-L using Ni-NTA resin columns. As shown in Fig. 7d (left and middle panels), PI3Kδ-S and -L purification was verified by western blotting using anti-His or anti-PI3Kδ antibody. Moreover, the Ni-NTA resin column approach resulted in the isolation of PI3Kδ isoforms that were no longer bound to p85α (Fig. 7d; far right panel, column E). Early reports have demonstrated that PI3Kα/p85 and PI3Kβ/p85 complexes are obligate and extremely stable, being able to withstand high concentrations of urea or detergent[42–44]. Our finding that PI3Kδ and p85α co-exist as monomers and complexes was unexpected. It should be noted, however, that the nature of interaction between PI3Kδ and p85 is less well established, as these two proteins have been shown in separate studies to either form a stable obligate complex[42] or coexist complexed together and uncomplexed from each other[45].

Purified PI3Kδ isoforms (minus p85α) were incubated with vehicle, nonselective PI3K inhibitor wortmannin (100 nM)[46] or PI3Kδ-specific inhibitor CAL-101 (100 nM), and subjected to a

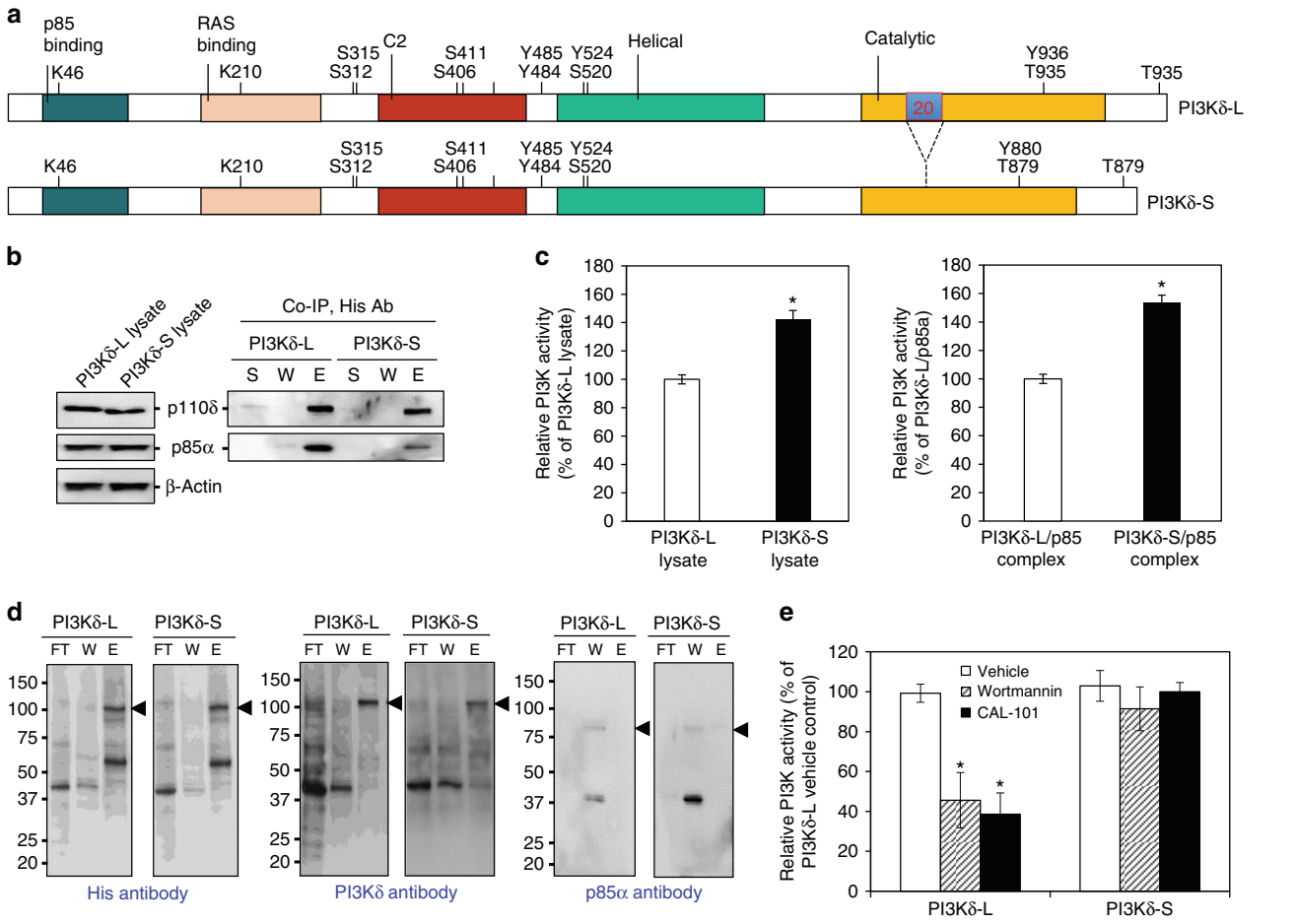

Figure 7 | Cell-free kinase assay of PI3Kδ isoforms and small-molecule inhibition. (**a**) Schematic representation of protein domains of PI3Kδ-L and -S isoforms. Adaptor (p85)-binding, RAS binding, C2, helix and catalytic domains are highlighted. Phosphorylation and ubiquitination sites (S, T, Y and K) and the region encoded by exon 20 (56 amino acids) residing in the catalytic domain are indicated. (**b**) Co-immunoprecipitation (Co-IP) of His-tagged PI3Kδ/p85α complex from transfected PC-3 cells followed by western blot, and (**c**) PI3K activity assays. S, supernatant; W, wash fraction; E, eluted fraction. Anti-His antibody was used in the Co-IP experiments, and anti-His, anti-p85α and anti-actin antibodies were used in the western blotting. *Significantly different kinase activities in total lysates of PI3Kδ-S versus PI3Kδ-L-expressing cells, or purified PI3Kδ-S/p85α versus PI3Kδ-L/p85α complexes. $P<0.05$ using two-sided Student's *t*-test. Data presented as mean ± s.e.m. of $n=4$ for each treatment group. (**d**) Purification of His-tagged PI3Kδ-L and -S isoforms. Western blot analysis of Ni-NTA resin-purified PI3Kδ isoforms from transfected PC-3 cells using His and PI3Kδ antibodies. FT, flow-through; W, wash fraction; E, eluted fraction. Closed arrowheads indicate PI3Kδ isoforms. (**e**) Cell-free kinase assay of L and S isoforms of PI3Kδ in the presence of vehicle (phosphate-buffered saline (PBS)), 100 nM wortmannin or 100 nM CAL-101. *Significantly different from vehicle control-treated PI3Kδ-L isoform. $P<0.05$ by ANOVA with Dunnett's *post hoc* test. Data presented as mean ± s.e.m. from at least four independent experiments for each treatment group. Blots in **b**,**d** are representative from at least three independent experiments with similar results. Variance was similar among groups being compared. Unprocessed western images are shown in Supplementary Fig. 13.

PI3K activity assay. In the absence of bound p85α, kinase activity of PI3Kδ-L was equivalent to PI3Kδ-S (Fig. 7e, compare vehicle treatments). In agreement, siRNA-mediated knockdown of p85α in wild-type VCaP and PC-3 cells was associated with an increase in invasive activity (Supplementary Fig. 8). Remarkably, wortmannin and CAL-101 significantly inhibited the activity of the PI3Kδ-L isoform, but not the PI3Kδ-S isoform (Fig. 7e). These results demonstrate that PI3Kδ-S maintains kinase activity even in the presence of SMIs, supporting the *in vitro* and *in vivo* results (Figs 5 and 6).

Discussion

The phenomenon of DS, much less global DS events, has not been adequately explored as a possible mechanism underlying PCa disparities. Potential involvement of the constitutively active *AR-V7* splice variant in PCa disparities has been suggested in a recent study. Selective downregulation of miR-212 observed in AA PCa is correlated with upregulation of splicing factor hnRNP-H1, upregulation of *AR-V7* and antiandrogen resistance in PCa cell lines[47]. In contrast to this localized splicing event, our study reveals that DS on a global scale may be a critical molecular mechanism underlying PCa disparities. In a comparison of AA PCa versus EA PCa, DS events were found to be highly prevalent in cancer-associated genes and pathways (Supplementary Data 1 and 2). Interestingly, the number of genes harbouring predicted DS events (2,520 genes) was ~3× greater than the number of differentially expressed, but not differentially spliced, genes (886 genes). These findings have two major implications. First, alternative/aberrant splicing of pre-mRNAs may have a greater role than differential gene expression in driving PCa disparities. Second, predicted DS events identified in our study were statistically overrepresented in oncogenic signalling pathways. In many cases, these same pathways are known to harbour a

preponderance of gene mutations across different cancer types[27–29]. Hence, DS adds another layer of complexity to the existing molecular repertoire (gene mutation, expression, methylation[7]) driving AA PCa aggressiveness.

Studies on AS indicate that approximately half of such events occurring in the coding sequence are in-frame, while the remaining events are frameshifts leading to truncated or extended C-terminal proteins[48,49]. Remarkably, 70% of AA-enriched/-specific variants in our composite oncogenic signalling pathway (Fig. 2) were in-frame, including *PIK3CD-S*, *FGFR2-S*, *FGFR3-S*, *TSC2-S*, *RASGRP2-b*, *ATM-S* and *GSK3-S* (Supplementary Table 2). In comparison, only 27.3% of EA-enriched/-specific DS events in our composite oncogenic signalling pathway exhibited in-frame preservation (Supplementary Table 2), while the remaining EA-enriched DS variants, including *ITGA4-S*, *MET-S*, *NF1-S*, *RASGRP2-a*, *mTOR-S* and *BAK1-S*, were frame-shifted. Why the vast majority of DS events appear to be in-frame in AA PCa versus frame-shifted in EA PCa remains unresolved (Supplementary Table 1). Presumably, the preponderance of AA in-frame events detected in oncogenic signalling pathways may be contributing to the more aggressive nature of AA PCa. Possible mechanisms that could drive differences in AS events include differential expression of *trans*-acting splicing factors[47] and/or single-nucleotide polymorphisms in *cis*-acting splicing elements of alternatively spliced genes[50]. In fact, a number of splicing factor mRNAs appear to be overexpressed (*SRSF2*, *SRSF7*) in AA PCa compared with EA PCa[9,10]. Regarding the in-frame variants (*PIK3CD-L*, *FGFR3-L* and *TSC2-L*) detected in EA PCa, each conferred a less aggressive oncogenic phenotype compared with the corresponding in-frame variants detected in AA PCa (*PIK3CD-S*, *FGFR3-S* and *TSC2-S*).

Approximately one-third of the AA-enriched/-specific variants identified in AA PCa were likewise present in patient-matched NP specimens, whereas the remaining AA-enriched/-specific variants found in PCa were absent in patient-matched NP specimens and thus appear to be *de novo* events (occurring as NP evolved into PCa). Accordingly, the AA-enriched/-specific variants already present in NP specimens have the potential to serve as inherent 'at-risk alleles' for poor PCa prognosis in AAs. In comparison, the *de novo* appearance of tumour-specific variants may drive poorer outcomes. *PIK3CD-S* would be an example of a potential AA 'at-risk allele' contributing to increased PCa aggressiveness upon disease presentation. Indeed, ectopic overexpression of the AA-enriched *PIK3CD-S* in PCa cell lines was demonstrated to enhance oncogenic potential (increased invasion, proliferation and AKT/mTOR signalling) compared with the corresponding EA-enriched *PIK3CD-L*. Moreover, genetic manipulation of AA MDA PCa 2b cells to favour expression of the *-S* variant over the *-L* variant likewise increased oncogenic behaviour. Conversely, genetic manipulation in the opposite direction decreased oncogenic behaviour. Interestingly, survival plots generated from The Cancer Genome Atlas (TCGA) RNA-sequencing data demonstrate that a high *S/L* ratio is associated with significantly worse survival for PCa and trending for worse survival in both breast and colon cancer (Supplementary Fig. 9). Survival plots were not stratified by race as this information is not currently available in TCGA. Given the number of patients analysed, it seems highly probable that high *S/L* ratio values may also be associated with a subset of EA patients, suggesting that *PIK3CD-S* may be useful in predicting survival in all patients irrespective of race. Besides *PIK3CD-S*, an additional 732 potential 'at-risk alleles' (for example, *ITGA4-L*, *MET-L*) were identified that may be associated with poor PCa prognosis in AAs. Further experimentation is needed to investigate whether these variants can serve as novel biomarkers to address PCa disparities. In contrast to the 'at-risk alleles', AA-enriched variants *FGFR-S* and *TSC2-S* were detected in AA PCa, but not in patient-matched NP specimens. The appearance of these *de novo* variants during PCa formation may contribute to driving the more aggressive PCa phenotype observed in the AA population, since *in vitro* genetic manipulation favoring expression of these *-S* variants over the *-L* variants promoted oncogenesis in MDA PCa 2b cells. It is noteworthy that three PCa-associated splice variants identified in previous studies, *Bcl-xL*, *FGFR2-IIIc* and *TMPRSS2-ERG + 72* (refs 14,15,20), did not exhibit DS in our AA PCa versus EA PCa comparison (Supplementary Data 1), suggesting that these variants may contribute to PCa progression/aggressiveness in a race-independent manner.

The identification of *PIK3CD-S*, a variant newly discovered and cloned in our study, as an 'at-risk allele' for PCa aggressiveness is germane given that PI3K signalling is aberrantly activated in a variety of cancers and PI3K inhibitors have been developed as targeted therapeutics[51,52]. Class IA PI3Ks consist of three isoforms, including PI3Kα, PI3Kβ and PI3Kδ. Unlike ubiquitously expressed PI3Kα and PI3Kβ, PI3Kδ appears to be preferentially expressed in leukocytes[53,54]. Previous studies have revealed a crucial role of PI3Kδ in lymphoid and myeloid malignancies[39,55]. Interestingly, accumulating evidence suggests a functional role of PI3Kδ in promoting nonhaematologic tumours as well. For example, overexpression of *PIK3CD* mRNA and/or PI3Kδ protein has been detected in glioblastoma[32], neuroblastoma[30], breast cancer[33] and PCa[31], and *PIK3CD* overexpression has been implicated in promoting cell growth/survival in breast cancer and neuroblastoma[30,33]. Consistent with these findings, our immunohistochemistry experiments using a pan-PI3Kδ antibody likewise revealed strong expression of PI3Kδ protein in PCa specimens as well as PCa, breast cancer and colon cancer cell lines (Supplementary Fig. 10). Importantly, our study provides greater granularity by being the first to demonstrate the relationship between expression of a race-enriched *PIK3CD* splice variant and cancer aggressiveness as well as resistance to SMIs targeting PI3Kδ.

Aberrant pre-mRNA splicing has recently been demonstrated to mediate therapeutic resistance in multiple cancer types. For example, the constitutively active *AR-V7* variant (lacking exonic sequences encoding the ligand binding domain) confers resistance to enzalutamide and abiraterone acetate in castration-resistant PCa patients[56]. In addition, melanoma patients harbouring *BRAF* splice variants encoding protein isoforms that are missing the RAS-binding domain exhibited resistance to the RAF inhibitor vemurafenib[48,49,57]. Noteworthy, these studies did not investigate whether variant expression and therapeutic responsiveness stratified along racial lines. We now provide evidence that AA-enriched *PIK3CD-S* imparts PCa cell lines with significant resistance to SMIs targeting PI3Kδ, as demonstrated in both *in vitro* assays and preclinical mouse models of PCa. This short variant is missing exon 20, encoding a 56-amino acid segment that is present in *PIK3CD-L*. Amino acids residing in the exon 20-encoded cassette appear critical for the docking of CAL-101 and wortmannin. Indeed, molecular modelling studies predict that Glu826 and Val828 (missing in PI3Kδ-S) undergo hydrogen bonding with CAL-101 (ref. 58). Noteworthy, overall response of indolent lymphoma and chronic lymphocytic leukaemia to CAL-101 ranges from 48 to 81% (refs 59–61). Given our findings, it would be of interest to determine whether patients with primary resistance harbour malignant cells expressing CAL-101-resistant PI3Kδ-S, while responsive patients harbour malignant cells expressing CAL-101-sensitive PI3Kδ-L.

P85 regulatory subunits are known binding partners of class I PI3Ks, resulting in protein stabilization and suppression of basal kinase activity[62,63]. Somatic mutations in *PIK3R1* (encoding

p85α) have been identified that abrogate the inhibitory action of p85α on PI3Kα in cancers[64,65]. Our cell-free assays demonstrated that p85α binds more efficiently with PI3Kδ-L compared with PI3Kδ-S. This interaction appears to be responsible for the lower kinase activity exhibited by PI3Kδ-L, as disruption of binding led to a long isoform with increased kinase activity comparable to PI3Kδ-S (Fig. 7c,e). The amino acid Asn334 located on the N-terminal side of PI3Kδ has been postulated to serve as a critical contact point with p85α (ref. 66). Our findings suggest that amino acids 810–865, encoded by exon 20 and missing in PI3Kδ-S, may also contain essential amino acids required for efficient coupling to p85α. Alternatively, amino acids 810–865 permits PI3Kδ-L to adopt a conformation where Asn334 (and other amino acids) is available to interact with p85α.

The identification and functional validation of global AS in cancer pathogenesis remains challenging and largely unexplored. We have undertaken such an analysis in the context of race-related aggressive PCa and identified a large number of DS events in cancer-associated pathways in EA and AA PCa, with a subset of these events also being detected in patient-matched NP specimens. These events will have both biological and clinical consequences, case in point *PIK3CD-S*. The identification of novel splice variants as biomarkers and/or development of therapeutics targeting protein isoforms have the potential to reduce cancer disparities.

Methods

Materials. EA PCa cell lines LNCaP (CRL-1740), VCaP (CRL-2876) and PC-3 (CRL-1435), and AA PCa cell line MDA PCa 2b (CRL-2422) were obtained from the American Type Tissue Collection (ATCC, Manassas, VA, USA), authenticated at the ATCC by short tandem repeat profiling of multiple unique genetic loci, and tested negative for mycoplasma. Primer sequences for RT-PCR are provided in Supplementary Table 3. The siRNAs were purchased from GE Dharmacon (Lafayette, CO, USA) and sequences are as follows: nonsense, 5′-CCA AAUUAU ACCUACAUUGCU-3′; siP_{20}, 5′-CCAACAUCCAACUCAACAA-3′; siP_j, 5′UGAGGGAGGCCCUGGAUCGA-3′; siF, 5′-CUCGACUACUACAAGAAGA-3′; siTSC2-ex20, CUGCGCUAUAAAGUGCUCA-3′; siRASGRP2-ex11, 5′-CCACAU CUCACAGGAAGAA-3′. siPIK3R1 Smart Pool (5′-AGUAAAGCAUUGUGUC AUA-3′, 5′-CCAACAACGGUAUGAAUAA-3′, 5′-GACGAGAGACCAAUAC UUG-3′, 5′-UAUUGAAGCUGUAGGGAA A-3′).

Collection of PCa clinical specimens. Prostate biopsy samples were collected at the George Washington University Medical Faculty Associates according to an institutional review board-approved protocol (IRB no. 020867). Informed consent was obtained from all study participants. High-quality PCa and patient-matched normal prostate (NP) biopsy cores from each of 20 AA and 15 EA primary PCa patients were collected and processed for the exon array analysis. PCa cores were determined by a pathologist to have Gleason scores of 6–7 (17 AA and 13 EA) or 8–9 (3 AA and 2 EA), while NP cores were diagnosed negative for cancer. There were no significant differences (*t*-test, $P>0.05$) between the two racial groups with respect to age (average age for AAs was 62.3 ± 8.2, average age for EAs was 63.3 ± 9.2) and Gleason score (range 6–8; Fisher's exact test, $P>0.05$). No distant metastasis was detected in the enrolled patients.

Exon array and statistical analyses. Total RNA was purified from PCa and patient-matched NP biopsy cores using the RNeasy micro kit as per manufacturer's protocol (Qiagen, Valencia, CA, USA). Briefly, total RNA samples were extracted using Trizol reagent, then treated with DNase I and further purified by RNeasy MinElute spin column. High-quality RNA isolation was confirmed by using the Agilent Bioanalyzer as per the manufacturer's protocol (Agilent Technologies, Santa Clara, CA, USA). For exon array analysis, 1 μg of purified RNA sample from each biopsy core was interrogated with the Affymetrix Human Exon 1.0 ST GeneChip (Santa Clara, CA, USA). Exon microarray data can be assessed at GEO (Gene Expression Omnibus) using accession number GSE64331. The exon array raw data were subjected to quantile normalization, GC-content adjustment, RMA background correction and $\log_2$ transformation using Partek Genomics Suite 6.6 software (Partek Incorporated, St Louis, MO, USA). Detection of differential expression at the gene level (gene-wise analysis) was performed in Partek using the One-Step Tukey's Biweight algorithm for detection of outlier probe-sets. Statistical analysis of exon expression data was based on ANOVA with multiple-correction testing using 10% false discovery rate (FDR)[67] criterion. DS events were modeled using the AS ANOVA algorithm[26] implemented in Partek together with selection of probe-sets exhibiting significant AS score determined at a 2% FDR. Principal component analysis plots and two-dimensional hierarchical clustering of exon-level data were performed using Partek. DS events were tested for statistical overrepresentation in canonical signalling pathways by Fisher's exact test using the Ingenuity Pathway Analysis (IPA) program (Ingenuity Systems, Redwood City, CA, USA).

RT-PCR validation of AS variants in AA and EA PCa. QRT-PCR was performed using the 7300 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) to validate and quantify AS events. Primers were designed to amplify the flanking regions of skipped exons or the junctions across catenated exons of variant mRNAs (Fig. 3b). Amplified RT-PCR products were quantified and normalized to house-keeping genes, *EIF1AX* and *PPA1*, using the ΔΔCt approach[9,10]. Primer sequences for RT-PCR validation are listed in Supplementary Table 3.

Molecular cloning of *PIK3CD-S* and *PIK3CD-L* variants and overexpression of variants in PCa cell lines. RT-PCR was performed to amplify *PIK3CD-L* and *PIK3CD-S* transcript variants from purified RNA of PC-3, VCaP and MDA PCa 2b cells (ATCC, Manassas, VA). PC-3 and VCaP cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Life Technologies, Gaithersburg, MD, USA) supplemented with 10% fetal bovine serum (FBS), while MDA PCa 2b cells were grown in BRFF-HPC-1 medium (AthenES, Baltimore, MD, USA) supplemented with 20% FBS. All the cell lines were grown at 37 °C and 5% CO_2. Primers were designed according to the National Center for Biotechnology Information (NCBI) reference sequences of *PIK3CD* mRNA (NM_005026.3). The forward primer contained the start codon (bold) (5′-**ATG**CCCCCTGGGGTGGACT-3′) and the reverse primer was upstream of the stop codon (5′-CTGCCTGTTGTCTTTG GACA-3′). Full-length PCR products were ligated into pcDNA3.1/V5-His TOPO vector (K4800-01, Invitrogen, Grand Island, NY, USA) using the manufacturer's protocol. A total of 8–10 independent clones were selected for each of the amplified *PIK3CD-L* and *PIK3CD-S* variants and sequence verified. The consensus sequences of *PIK3CD-S* and *PIK3CD-L* mRNAs were deposited to GeneBank (accession number KU612116 and KU612117). The plasmids pcDNA3.1-PIK3CD-L/V5-His and pcDNA3.1-PIK3CD-S/V5-His were individually transfected into the PCa cell lines (VCaP and PC-3) using the cationic lipid-mediated method[9] to establish stable cell lines overexpressing *PIK3CD-L* or *PIK3CD-S*.

Functional analysis of PCa cell lines following transfection of siRNAs targeting the *PIK3CD-S* or the *PIK3CD-L* variant. VCaP and MDA PCa 2b cells were grown in DMEM with 10% FBS for 24 h, and then were transfected for 24 h with siRNAs (50 nM) designed to target splice variants of *PIK3CD*, *FGFR3*, *TSC2* or *RASGRP2* using DharmaFECT4 transfection reagent (Dharmacon), according to the manufacturer's protocol. The *in vitro* functional assays, including cell proliferation and invasion, were performed following siRNA transfections for 24 h. Cell proliferation and invasion assays were performed using 5-bromodeoxyuridine Cell Proliferation Assay kit (Calbiochem, Billerica, MA, USA) and the Matrigel Invasion Chambers (BD Biosciences, San Jose, CA, USA), respectively, as per the manufacturers' protocol[9,10].

Antibodies. Antibodies used in western blot analysis were rabbit monoclonal antibodies for $pAKT^{Tyr308}$, $pAKT^{Ser473}$, AKT, pmTOR, mTOR, pS6 and S6 (2965, 4058, 4691, 2971, 2983, 4857 and 2983, Cell Signaling Technology, Danvers, MA, USA), rabbit polyclonal antibodies for His-tag (ab18184, Abcam, Cambridge, MA, USA), HA-tag, PI3Kδ, p85α and β-actin (sc-7392, sc-55589, sc-1637 and sc-4778, Santa Cruz Biotechnology, Santa Cruz, CA, USA). Horseradish peroxidase-conjugated secondary antibodies for rabbit and mouse IgG were purchased from Southern Biotech (Birmingham, AL, USA).

***In vivo* xenograft and metastasis models.** All animal work was approved by the George Washington University institutional animal care and use committee (protocol A272). Male NOD-SCID mice, 4–6 weeks old, were purchased from the Jackson Laboratory (Bar Harbor, ME, USA). To establish a PCa xenograft model, 2×10^6 PC-3 cells stably overexpressing *PIK3CD-L* or *PIK3CD-S* were subcutaneously injected into the left flank of NOD-SCID mice. Tumour xenograft growth was measured with calipers and the volume was determined as $1/2\times$ length $\times$ width^2. Mice were randomized into groups once the average tumour size reached $\sim200\,mm^3$ and treated with vehicle (phosphate-buffered saline) or CAL-101 ($50\,mg\,kg^{-1}$) through daily i.p. injections. After 30 days, mice were killed and the dissected xenografts were photographed and weighed using a blinded design.

To establish the PCa metastasis model, 1×10^6 PC-3 cells stably overexpressing *PIK3CD-L* or *PIK3CD-S* were injected into the tail vein of NOD-SCID mice. The mice were then treated with vehicle or CAL-101 ($50\,mg\,kg^{-1}$) via i.p. injections, 3 times a week. After 8 weeks, lungs of mice were collected and stained with India ink and bleached with Fekete's solution (70% ethanol, 3.7% formaldehyde, 0.75 M glacial acetic acid). India ink-stained lungs were photographed and lung metastases were quantified using the NIH ImageJ program[68].

Purification of His-tagged PI3Kδ protein. PC-3 cells stably overexpressing *PIK3CD-L* or *PIK3CD-S* were maintained in DMEM (Life Technologies) supplemented with 10% FBS. After growing the cells for 24 h, cell extracts were prepared and His-tagged PI3Kδ protein was purified using a column HisPur Ni-NTA purification kit (Pierce Biotechnology, Rockford, IL, USA). Briefly, cell lysates were mixed with Ni-NTA resin and incubated at room temperature for 30 min. After incubation, the resin was washed with wash buffer (25 mM imidazole, pH 7.4) and applied to a HisPur Ni-NTA spin column, centrifuged and wash buffer eluate discarded after centrifugations. His-tagged proteins were eluted from the resin by adding one-resin-bed volume of elution buffer (250 mM imidazole, pH 7.4). The purified PI3Kδ-His protein was mixed with 2 × Laemmli sample buffer, boiled and analysed by immunoblotting.

Co-IP of PI3Kδ/p85 complex. Plasmids pcDNA3.1-PIK3CD-S/V5-His (or pcDNA3.1-PIK3CD-L/V5-His) and pSV-p85α (Addgene, Cambridge, MA, USA) were co-transfected into PC-3 cells. After growing the cells for 48 h, the co-transfected cells were collected and cells were lysed with RIPA lysis buffer (Santa Cruz Biotechnology). The cell lysates were then subjected to Co-IP assays with anti-His antibody (ab18184, Abcam) and immobilized on protein G-Sepharose beads (Thermo Scientific, Waltham, MA, USA). Cell lysates and precipitates were subjected to western blotting, and visualized by enhanced chemiluminescence system (Thermo Scientific, Waltham, MA, USA).

***In vitro* assay of PI3Kδ activity.** PI3Kδ activity was evaluated with a PI3K activity/inhibitor assay kit (Millipore, Billerica, MA, USA) according to the manufacturer's instructions. Briefly, purified His-tagged PI3Kδ-L or PI3Kδ-S isoform was pretreated with the PI3Kδ inhibitor (100 nM of wortmannin or 100 nM of CAL-101) or vehicle in 96-well plates for 10 min and subjected to a competitive ELISA. PIP2 substrate and kinase reaction buffer were added to the pretreated His-tagged PI3Kδ-L or PI3Kδ-S isoform and incubated at room temperature for 1 h. After incubation, biotinylated PIP3 and GST-GRP1 working solutions were added to the wells and the reaction samples were further incubated at room temperature for 1 h. Plates were washed three times with 1 × Tris-buffered saline with Tween-20 (50 mM Tris-Cl, pH 7.5. 150 mM NaCl) and incubated with streptavidin-horseradish peroxidase conjugate (1.25 mg ml^{-1}) at room temperature for 1 h. After incubation, plates were washed three times and incubated with 100 µl of TMB (3,3′,5,5′-tetramethylbenzidine, 1 mg ml^{-1}) substrate solution at room temperature for 5–20 min. Reactions were stopped by adding 100 µl of stop solution and plates read at 450 nm. The colourimetric signal was inversely proportional to the amount of PIP3 produced by PI3K activity and the relative amount of PIP3 produced was determined with a standard curve.

Analysis of in-frame and out-of-frame exon skipping events. A total of 4,253 significant differentially expressed exons (2% FDR) were identified. Using a 1.5-fold cutoff, the number of differentially expressed exons (that is, exon skipping events) was narrowed down to 3,112 (corresponding to 2,520 DS genes in AA PCa versus EA PCa). The reference coordinates for the Affymetrix probe sets used to identify these skipped exons were cross-referenced with the Ensembl database, release GRCh37, resulting in final set of 2,517 well-curated, differentially expressed exons that corresponded to 1,484 DS genes (Supplementary Data 4). Based on the exon size and the modelled effect, the exon skipping events were classified as in-frame or frame-shift (Supplementary Data 4). A total of 1376 genes were included in this analysis. The distribution of in-frame and frameshift events was then compared between the EA and AA groups. We grouped the observations based on presence of at least one frameshift per gene. Notably, we observed significantly higher proportion of alternatively spliced genes without a frameshift in the AA group (34% in AA versus 27% in EA, $P<0.005$, Fisher's exact test).

Metrics for assessing reliability of global analysis of differential splicing events. Affymetrix Exon GeneChip analysis has revealed ~2,500 differential splicing events between AA and EA PCa. While we can only draw firm conclusions on a subset of differential splicing patterns that were validated by a second approach (that is, RT-PCR), the following metrics have been provided to allow evaluation of the overall reliability of Exon Genechip results:

(1) *Validation success rate*. Eight of 9 genes (89%) identified by Exon GeneChip analysis to undergo differential RNA splicing between AA PCa and EA PCa were successfully validated by quantitative RT-PCR. The only gene that did not validate was *EPHA1*. In addition, 2 out of 2 genes (100%), defined not to exhibit differential RNA splicing (*GSK3A* and *ATM*), were successfully validated by quantitative RT-PCR. Aggregate success rate was 91% (10/11).

(2) *P values*. The *P* values ranged from 1×10^{-3} to 1×10^{-15} for the 9 genes identified by Exon GeneChip analysis to undergo differential RNA splicing between AA and EA PCa specimens ($P=6\times10^{-4}$ for *ATM*). In comparison, *P* values for all genes identified by Exon GeneChip analysis to exhibit differential RNA splicing ranged from 3×10^{-3} to 1×10^{-20} (see Supplementary Table 1). Hence, *P* values of genes chosen for RT-PCR validation were representative of the entire range of *P* values associated with the complete set of ~2,500 genes identified by Exon ChipGene analysis.

(3) *Power calculations*. Based on the number of PCa patient samples interrogated by our Affymetrix Exon GeneChip arrays ($n=15$ to 20 patients per gene per race) and a computed s.d. = 0.4, our findings correspond to >85% power to distinguish 1.5-fold changes at $P<0.01$.

Estimation of *PIK3CD* isoform expression. Expression of the short and long isoforms of *PIK3CD* was determined using the method IsoformEx[69]. Briefly, isoform expression was estimated through the minimizing a weighted nonnegative least squares problem based on the exon expression. For the purpose of this analysis, the novel 'short isoform' was defined as any transcript missing exon 20 but having exons 19 and 21 concatenated; and the 'long isoforms' were defined as any transcripts with exons 19, 20 and 21 concatenated. Raw data for breast invasive carcinoma (BRCA), prostate adenocarcinoma (PRAD) and colorectal adenocarcinoma (COAD) were obtained from TCGA RNA-sequencing exon expression (https://tcga-data.nci.nih.gov/tcga/, accessed 22 January 2016). The ratio of short to long isoforms was calculated for survival analysis.

Survival analysis. Disease-free survival data for BRCA, COAD and PRAD were obtained from TCGA clinical data (https://tcga-data.nci.nih.gov/tcga/, accessed 22 January 2016). Patients who did not have a relapse event during the study were considered as censored. The expression values of the short and long isoforms as well as the interaction term were used as predictors to fit the Cox proportional hazards regression model under L2-regularization, where the disease-free survival is the response variable. For each patient, a prognosis index score was computed from the Cox proportional hazards model[70]. Briefly, the patients were dichotomized into high- and low-risk groups according to the relapsed versus relapse-free ratio. The log-rank *P* value was then calculated to assess the statistically significant difference between the Kaplan–Meier curves of the high- versus low-risk groups.

Data availability. The authors declare that all data supporting the findings of this study are available within the article and its Supplementary Information files or from the corresponding author on reasonable request.

References

1. Jemal, A. *et al.* Cancer statistics, 2007. *CA Cancer. J. Clin.* **57,** 43–66 (2007).
2. Reddy, S., Shapiro, M., Morton, Jr R. & Brawley, O. W. Prostate cancer in black and white Americans. *Cancer Metastasis. Rev.* **22,** 83–86 (2003).
3. Powell, I. J. Epidemiology and pathophysiology of prostate cancer in African-American men. *J. Urol.* **177,** 444–449 (2007).
4. Jones, B. A. *et al.* Explaining the race difference in prostate cancer stage at diagnosis. *Cancer Epidemiol. Biomarkers Prev.* **17,** 2825–2834 (2008).
5. Evans, S., Metcalfe, C., Ibrahim, F., Persad, R. & Ben-Shlomo, Y. Investigating Black-White differences in prostate cancer prognosis: a systematic review and meta-analysis. *Int. J. Cancer.* **123,** 430–435 (2008).
6. Kittles, R. A. *et al.* A common nonsense mutation in EphB2 is associated with prostate cancer risk in African American men with a positive family history. *J. Med. Genet.* **43,** 507–511 (2006).
7. Powell, I. J. & Bollig-Fischer, A. Minireview: the molecular and genomic basis for prostate cancer health disparities. *Mol. Endocrinol.* **27,** 879–891 (2013).
8. Wallace, T. A. *et al.* Tumor immunobiological differences in prostate cancer between African-American and European-American men. *Cancer Res.* **68,** 927–936 (2008).
9. Wang, B. D. *et al.* Identification and functional validation of reciprocal microRNA-mRNA pairings in African American prostate cancer disparities. *Clin. Cancer Res.* **21,** 4970–4984 (2015).
10. Wang, B. D. *et al.* Androgen receptor-target genes in african american prostate cancer disparities. *Prostate Cancer* **2013,** 763569 (2013).
11. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40,** 1413–1415 (2008).
12. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24,** 2343–2364 (2010).
13. Chen, J. & Weiss, W. A. Alternative splicing in cancer: implications for biology and therapy. *Oncogene* **34,** 1–14 (2015).
14. Wang, J. *et al.* Pleiotropic biological activities of alternatively spliced TMPRSS2/ERG fusion gene transcripts. *Cancer Res.* **68,** 8516–8524 (2008).
15. Mercatante, D. R., Mohler, J. L. & Kole, R. Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. *J. Biol. Chem.* **277,** 49374–49382 (2002).
16. Rajan, P., Elliott, D. J., Robson, C. N. & Leung, H. Y. Alternative splicing and biological heterogeneity in prostate cancer. *Nat. Rev. Urol.* **6,** 454–460 (2009).
17. Akgul, C., Moulding, D. A. & Edwards, S. W. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell. Mol. Life Sci.* **61,** 2189–2199 (2004).
18. Sun, A. *et al.* Androgen receptor-dependent regulation of Bcl-xL expression: Implication in prostate cancer progression. *Prostate* **68,** 453–461 (2008).

19. Bauman, J. A., Li, S. D., Yang, A., Huang, L. & Kole, R. Anti-tumor activity of splice-switching oligonucleotides. *Nucleic Acids Res.* **38,** 8348–8356 (2010).
20. Oltean, S. *et al.* Alternative inclusion of fibroblast growth factor receptor 2 exon IIIc in Dunning prostate tumors reveals unexpected epithelial mesenchymal plasticity. *Proc. Natl Acad. Sci. USA* **103,** 14116–14121 (2006).
21. Demichelis, F. *et al.* TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* **26,** 4596–4599 (2007).
22. Hu, Y. *et al.* Delineation of TMPRSS2-ERG splice variants in prostate cancer. *Clin. Cancer Res.* **14,** 4719–4725 (2008).
23. Heinlein, C. A. & Chang, C. Androgen receptor in prostate cancer. *Endocr. Rev.* **25,** 276–308 (2004).
24. Antonarakis, E. S. *et al.* AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N. Engl. J. Med.* **371,** 1028–1038 (2014).
25. Hu, R. *et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.* **69,** 16–22 (2009).
26. Fan, W., Stirewalt, D. L., Radich, J. P. & Zhao, L. Comparison of two methods for detecting alternative splice variants using genechip((R)) exon arrays. *Int. J. Biomed. Sci.* **7,** 172–180 (2011).
27. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
28. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455,** 1069–1075 (2008).
29. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2012).
30. Boller, D. *et al.* Targeting the phosphoinositide 3-kinase isoform p110delta impairs growth and survival in neuroblastoma cells. *Clin. Cancer Res.* **14,** 1172–1181 (2008).
31. Jiang, X., Chen, S., Asara, J. M. & Balk, S. P. Phosphoinositide 3-kinase pathway activation in phosphate and tensin homolog (PTEN)-deficient prostate cancer cells is independent of receptor tyrosine kinases and mediated by the p110beta and p110delta catalytic subunits. *J. Biol. Chem.* **285,** 14980–14989 (2010).
32. Knobbe, C. B. & Reifenberger, G. Genetic alterations and aberrant expression of genes related to the phosphatidyl-inositol-3′-kinase/protein kinase B (Akt) signal transduction pathway in glioblastomas. *Brain Pathol.* **13,** 507–518 (2003).
33. Tzenaki, N. *et al.* High levels of p110delta PI3K expression in solid tumor cells suppress PTEN activity, generating cellular sensitivity to p110delta inhibitors through PTEN activation. *FASEB J.* **26,** 2498–2508 (2012).
34. Glickman, M., Malek, R. L., Kwitek-Black, A. E., Jacob, H. J. & Lee, N. H. Molecular cloning, tissue-specific expression, and chromosomal localization of a novel nerve growth factor-regulated G-protein- coupled receptor, nrg-1. *Mol. Cell Neurosci.* **14,** 141–152 (1999).
35. Korenchuk, S. *et al.* VCaP, a cell-based model system of human prostate cancer. *In Vivo* **15,** 163–168 (2001).
36. Navone, N. M. *et al.* Establishment of two human prostate cancer cell lines derived from a single bone metastasis. *Clin. Cancer Res.* **3,** 2493–2500 (1997).
37. Herman, S. E. *et al.* Phosphatidylinositol 3-kinase-delta inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. *Blood* **116,** 2078–2088 (2010).
38. Lannutti, B. J. *et al.* CAL-101, a p110delta selective phosphatidylinositol-3-kinase inhibitor for the treatment of B-cell malignancies, inhibits PI3K signaling and cellular viability. *Blood* **117,** 591–594 (2011).
39. Fruman, D. A. & Rommel, C. PI3Kdelta inhibitors in cancer: rationale and serendipity merge in the clinic. *Cancer Discov.* **1,** 562–572 (2011).
40. Yang, J. *et al.* Pathological axonal death through a MAPK cascade that triggers a local energy deficit. *Cell* **160,** 161–176 (2015).
41. Berndt, A. *et al.* The p110 delta structure: mechanisms for selectivity and potency of new PI(3)K inhibitors. *Nat. Chem. Biol.* **6,** 117–124 (2010).
42. Geering, B., Cutillas, P. R., Nock, G., Gharbi, S. I. & Vanhaesebroeck, B. Class IA phosphoinositide 3-kinases are obligate p85-p110 heterodimers. *Proc. Natl Acad. Sci. USA* **104,** 7809–7814 (2007).
43. Fry, M. J. *et al.* Purification and characterization of a phosphatidylinositol 3-kinase complex from bovine brain by using phosphopeptide affinity columns. *Biochem. J.* **288 Pt 2** 383–393 (1992).
44. Woscholski, R., Dhand, R., Fry, M. J., Waterfield, M. D. & Parker, P. J. Biochemical characterization of the free catalytic p110 alpha and the complexed heterodimeric p110 alpha.p85 alpha forms of the mammalian phosphatidylinositol 3-kinase. *J. Biol. Chem.* **269,** 25067–25072 (1994).
45. Fransson, S. *et al.* p37delta is a new isoform of PI3K p110delta that increases cell proliferation and is overexpressed in tumors. *Oncogene* **31,** 3277–3286 (2012).
46. Arcaro, A. & Wymann, M. P. Wortmannin is a potent phosphatidylinositol 3-kinase inhibitor: the role of phosphatidylinositol 3,4,5-trisphosphate in neutrophil responses. *Biochem. J.* **296 Pt 2** 297–301 (1993).
47. Yang, Y. *et al.* Dysregulation of microRNA-212 promotes castration resistance via hnRNPH1-mediated regulation of AR and AR-V7: implications for racial disparity of prostate cancer. *Clin. Cancer Res.* (2015).
48. Shi, H. *et al.* Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov.* **4,** 80–93 (2014).
49. Wellbrock, C., Karasarides, M. & Marais, R. The RAF proteins take centre stage. *Nat. Rev. Mol. Cell Biol.* **5,** 875–885 (2004).
50. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126,** 37–47 (2006).
51. Courtney, K. D., Corcoran, R. B. & Engelman, J. A. The PI3K pathway as drug target in human cancer. *J. Clin. Oncol.* **28,** 1075–1083 (2010).
52. Ciraolo, E., Morello, F. & Hirsch, E. Present and future of PI3K pathway inhibition in cancer: perspectives and limitations. *Curr. Med. Chem.* **18,** 2674–2685 (2011).
53. Clayton, E. *et al.* A crucial role for the p110delta subunit of phosphatidylinositol 3-kinase in B cell development and activation. *J. Exp. Med.* **196,** 753–763 (2002).
54. Jou, S. T. *et al.* Essential, nonredundant role for the phosphoinositide 3-kinase p110delta in signaling by the B-cell receptor complex. *Mol. Cell. Biol.* **22,** 8580–8591 (2002).
55. Castillo, J. J., Furman, M. & Winer, E. S. CAL-101: a phosphatidylinositol-3-kinase p110-delta inhibitor for the treatment of lymphoid malignancies. *Expert Opin. Investig. Drugs* **21,** 15–22 (2012).
56. Nakazawa, M., Antonarakis, E. S. & Luo, J. Androgen receptor splice variants in the era of enzalutamide and abiraterone. *Horm. Cancer* **5,** 265–273 (2014).
57. Freeman, A. K., Ritt, D. A. & Morrison, D. K. The importance of Raf dimerization in cell signaling. *Small GTPases* **4,** 180–185 (2013).
58. Berndt, A. *et al.* The p110delta structure: mechanisms for selectivity and potency of new PI(3)K inhibitors. *Nat. Chem. Biol.* **6,** 117–124 (2010).
59. Gopal, A. K. *et al.* PI3Kdelta inhibition by idelalisib in patients with relapsed indolent lymphoma. *N. Engl. J. Med.* **370,** 1008–1018 (2014).
60. Brown, J. R. *et al.* Idelalisib, an inhibitor of phosphatidylinositol 3-kinase p110delta, for relapsed/refractory chronic lymphocytic leukemia. *Blood* **123,** 3390–3397 (2014).
61. Shah, A. & Mangaonkar, A. Idelalisib: a novel PI3Kdelta inhibitor for chronic lymphocytic leukemia. *Ann. Pharmacother.* **49,** 1162–1170 (2015).
62. Vanhaesebroeck, B., Guillermet-Guibert, J., Graupera, M. & Bilanges, B. The emerging mechanisms of isoform-specific PI3K signalling. *Nat. Rev. Mol. Cell Biol.* **11,** 329–341 (2010).
63. Yu, J. *et al.* Regulation of the p85/p110 phosphatidylinositol 3′-kinase: stabilization and inhibition of the p110alpha catalytic subunit by the p85 regulatory subunit. *Mol. Cell Biol.* **18,** 1379–1387 (1998).
64. Chan, T. O. *et al.* Small GTPases and tyrosine kinases coregulate a molecular switch in the phosphoinositide 3-kinase regulatory subunit. *Cancer Cell* **1,** 181–191 (2002).
65. Jaiswal, B. S. *et al.* Somatic mutations in p85alpha promote tumorigenesis through class IA PI3K activation. *Cancer Cell* **16,** 463–474 (2009).
66. Amzel, L. M. *et al.* Structural comparisons of class I phosphoinositide 3-kinases. *Nat. Rev. Cancer* **8,** 665–669 (2008).
67. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* **57,** 289–300 (1995).
68. Fukuda, H. *et al.* Host-derived MMP-13 exhibits a protective role in lung metastasis of melanoma cells by local endostatin production. *Br. J. Cancer* **105,** 1615–1624 (2011).
69. Kim, H., Bi, Y., Pal, S., Gupta, R. & Davuluri, R. V. IsoformEx: isoform level gene expression estimation using weighted non-negative least squares from mRNA-Seq data. *BMC Bioinformatics* **12,** 305 (2011).
70. Huang, S., Yee, C., Ching, T., Yu, H. & Garmire, L. X. A novel model to combine clinical and pathway-based transcriptomic information for the prognosis prediction of breast cancer. *PLoS Comput. Biol.* **10,** e1003851 (2014).

Acknowledgements

This research was supported by ACS IRG-08-091-01 Grant (to B.-D.W.), NCI Grant U01CA11637-05S2 (to S.R.P.) and Affymetrix Collaborations in Cancer Research Award, DOD Grant PCRP:PC121975 and NCI Grant CA204806 (to N.H.L.).

Author contributions

B.-D.W. and N.H.L. conceived and designed the study. B.-D.W. and J.O. performed the microarray experiments, tissue sample analysis and *in vitro* assays. B.-D.W., V.P., K.C. and S.H. were involved in animal work. R.A. collected clinical samples for the study. B.-D.W., N.H.L., A.H., T.C., L.G. and M.A.G.-B. carried out genomic data analysis. B.-D.W. and S.K. performed three-dimensional (3D) modelling of protein isoforms. S.R.P. initiated and organized clinical sample collection at beginning of study and provided input in the Discussion section. B.-D.W., J.A.F. and N.H.L. wrote the manuscript with input from all authors. N.H.L. supervised the project.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wang, B.-D. *et al.* Alternative splicing promotes tumor aggressiveness and drug resistance in African American prostate cancer. *Nat. Commun.* **8**, 15921 doi: 10.1038/ncomms15921 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.